The growth effect of disaggregated foreign aid : evidence from bilateral loans and grants

Danny Kurban

The Growth Effect of Disaggregated Foreign Aid: Evidence from Bilateral Loans and Grants

Danny Kurban

Université Paris 1 Panthéon-Sorbonne

UFR 02 – Sciences Économiques

Master 2 Recherche Économie de la Mondialisation

Supervisor: Lisa Chauvet

June 3rd, 2015

L'université de Paris 1 Panthéon-Sorbonne n'entend donner aucune approbation ni désapprobation aux opinions émises dans ce mémoire: elles doivent être considérées comme propre à leur auteur.

The University of Paris 1 Panthéon-Sorbonne neither approves nor disapproves of the opinions expressed in this dissertation: they should be considered as the author's own.

Contents

1. Introduction

The share devoted to foreign aid in developed countries' government budgets is usually small. But on the side of the recipient country, aid inflows can be rather large, as a share of government revenues and also relative to GDP. Foreign aid thus provides the opportunity for donors to obtain a large positive impact on living conditions in developing countries with relatively little financial effort. However, whether aid really has a positive and measurable effect on economic growth and other development indicators has been intensely debated by economists, with no scientific consensus reached so far. Arndt, Jones, and Tarp (2014) claim that in the long run aid has increased growth, as well as life expectancy and the average years of schooling, while decreasing poverty and infant mortality. Doucouliagos and Paldam (2011) on the other hand show in a meta-study of 105 aid-growth studies that on average, and if the publication bias is controlled for, no significant effect of foreign aid on economic growth has been found in the literature. They further suggest to shift the focus away from aggregate aid measures to more disaggregated ones, which is one of the main motivations for this study. More specifically, the present study disaggregates total aid flows into its grant and loan components and asks if at the least one of them has a significant impact on growth. If both are significant, the question becomes whether their effects are significantly different *from each other.*

Much like in the case of aid effectiveness in general, the question whether to give foreign aid as non-repayable grants or as repayable loans has been the issue of an intense debate. However, unlike in the aid-growth debate, the discussion has been driven by political rather than economic arguments. From the donor perspective, the decision whether to give aid as grants or loans is probably more motivated by political considerations rather than by their relative efficiency (Sanford (2002), Nunnenkamp, Thiele, and Wilfer (2005)). Nevertheless, there is a strand of literature that assesses the different economic implications of grants and loans with respect to different outcomes, which may be substantial. But again, contrary to the general aid and growth debate, this literature has remained largely theoretical. Empirical results on the differential effects of loan and grant aid in the receiving countries are scarce, perhaps due to the various pitfalls and difficulties that arise if one aims to identify them. Identifying a causal effect of aggregated aid on growth is already challenging, due to the widely recognized problem of endogeneity bias. This has resulted in a large array of studies developing methods to overcome this bias. But if the goal is to empirically identify the impact of disaggregated aid, that is grants and loans, separately, the econometric challenges become even greater, due to the additional problem of multicollinearity and joint determination. Nevertheless, the theoretical arguments for different effects of loans and grants seem strong enough to warrant a comprehensive empirical investigation, despite the obvious difficulties lying ahead.

Hence, the research question addressed in this paper asks about the differential effect of two highly correlated and most likely endogenous factors on growth. Even though there are reasons to believe that the *true* effects of grants and loans differ, as the next section will show, this difference may be so small that it is practically impossible to detect and economically insignificant. However, the failure to detect any significant differential effect in the empirical estimation does not mean that it is too small to be detected or even non-existent. It could also be due to the problems of multicollinearity and weak instrumentation, especially in the case of multiple endogenous regressors. If one endogenous regressor gets insignificant when instrumented, it becomes practically impossible to identify whether it is significantly different from other regressors, because confidence bands usually overlap. Thus, in light of these problems, a second contribution of this study, in addition to trying to answer the research question about the growth effects of grants and loans in a satisfactory way, is of a more methodological nature. Using the example of aid-growth regressions with multiple endogenous regressors, the problems of weak instrumentation and multicollinearity as well as the most effective and feasible ways to detect and address them will be discussed. This is especially the case for Sections 4 and 5, which aim at analyzing these problems in an applied manner.

Section 2 gives an overview of the general debate about the effects of aid on growth and the related literature, followed by a summary of the theoretical considerations in favor of loans and grants, respectively. Section 3 introduces the data and shows a few key summary statistics. In Section 4, a detailed explanation of the endogeneity problem in the aid literature and its consequences for empirical studies is given. Afterwards, four of the main identification strategies developed to deal with this problem are analyzed, with a special focus on their feasibility in the context of the specific research question addressed in this paper. Section 5 applies modified versions of two of these strategies and tries to identify an effect of aid, disaggregated to grants and loans, on growth. Section 6 shows the results of an extensive set of robustness tests that have been run in order to address some major threats to the validity of the results, and Section 7 concludes.

2. Theory and literature

This chapter briefly discusses how foreign aid can affect GDP growth from a theoretical point of view and then gives an overview of the previous empirical studies and their most important results (Section 2.1). Due to the vast amount of literature on the topic, this overview remains selective in the choice of works and rather short in their treatment. However, some of the most influential methodological papers will be discussed in greater detail in Section 4. Section 2.2 then reviews the theoretical and empirical contributions that have already been made with regard to the specific research question addressed in this paper: "Is there a difference in the effectiveness of grant aid and loan aid?"

2.1 On the heterogeneous effect of foreign aid on growth

In a standard neoclassical growth model, capital accumulation (*i.e.* investment) is an important determinant of GDP growth. However, especially for less developed countries, investment is constrained by the difficult access to external sources of financing, as these countries face particularly high borrowing costs on international capital markets. This borrowing constraint can lead to savings and investment rates which are too low from an intertemporal point of view, holding the growth rate below its optimal level. Foreign aid, whether it comes as a permanent transfer (grant) or as a temporal one (concessional loans), can relax the borrowing constraint and thus promote growth. However, in the long run this will only happen when at least part of the aid transfer is used for productive investment, as opposed to unproductive investment or consumption. It is also important that aid is not highly fungible, meaning that it does not just crowd out domestically financed investments and in the process lead to reallocation of resources to other, less productive government activities. Only if aid finances productive investments and if fungibility is low will it help to bring a country closer to its optimal steady state long-run growth rate. In other words, the effect of foreign aid on growth is of course conditional on its use, which in turn can depend on many factors, such as the specific type of aid (technical assistance, grants, loans etc.) or the policy environment in the receiving country. As a result, the effect of aid on growth is expected to be heterogenous across recipient countries and over time.

This theoretical conjecture has been tested in various ways, resulting in a large body of empirical literature.[1] The study of Boone (1996) can be seen as the starting point of this “modern” strand of literature which considers a possibly heterogenous impact of aid. Boone finds that aid has increased neither investment nor different human development indicators, such as infant mortality and life expectancy. Furthermore, aid ineffectiveness seems to be independent of the political regime. He explains this non-significant impact of aid with the fact that politicians who maximize their own welfare use aid mainly to increase consumption instead of financing productive investments. These findings are in a way consistent with the much more influential results of Burnside and Dollar (2000), who find in their seminal paper that aid has an insignificant effect on growth, unless it is combined with good policies.[2] In their framework, this is shown by interacting a policy index, combining the inflation rate, the budget surplus, and a trade openness index, with the aid flow variable. The coefficient of this interaction term turns out to be positive and highly significant, which underlines their main conclusion that aid is effective only when it is combined with good policies.

While in the aftermath many scholars used similar empirical strategies to test for the conditional effect of aid on growth with respect to many different variables,[3] the result of Burnside and Dollar has also been criticized by others. Among the critics are Hansen and Tarp (2001), who show that the aid-policy interaction term loses significance if another possible non-linearity of aid is accounted for. More specific, they add a squared aid variable to the specification of Burnside and Dollar, which turns out to have a negative and significant coefficient, suggesting that aid affects growth only with diminishing marginal returns, independent of the policy regime. Easterly (2003) shows that the Burnside and Dollar results are very sensitive with respect to alternative definitions of their main variables and he also criticizes the way in which the “aid industry” has used these results. In a follow-up, Easterly, Levine, and Roodman (2004) show that the Burnside and Dollar results are not robust to an expansion of the

[1] A good synthesis on the heterogenous impact of aid on growth is given by Chauvet (forthcoming).

[2] Unlike Burnside and Dollar (2000), Boone (1996) did not look at the effectiveness of aid conditional on actual policy outcomes, but rather conditional on the institutional environment („liberal democracy“ *versus* „repressive regime“).

[3] For example, Dalgaard, Hansen, and Tarp (2004) find that aid is less effective in the tropics, Djankov, Montalvo, and Reynal-Querol (2009) show that aid is less effective when it comes from too many different donors and Lessmann and Markwardt (2012) find that aid is more effective in highly centralized economies. More recently, Dreher, Eichenauer, and Gehring (2014) and Dreher, Minasyan, and Nunnenkamp (2014) looked at the effect of political variables at the donor and donor-recipient levels and found that aid given for political (strategic) reasons is less effective.

dataset with respect to the country sample and the time frame considered. Guillaumont and Chauvet (2001) find that the external and climate environment of an aid receiving country matters more for the effectiveness of aid than good policies. Using a novel identification strategy, in line with Tavares (2003), Rajan and Subramanian (2008) find no significant impact of aid on growth, which leads them to conclude that previous studies have been plagued by endogeneity bias. However, Arndt, Jones, and Tarp (2010) show that a slightly improved, but very similar instrumentation strategy again confirms a positive and unconditional effect of aid on growth, at least over longer time frames. A summary of the often contradictory results on this topic is given in the extensive meta-studies of the aid-growth literature by Doucouliagos and Paldam (2008, 2009, 2011), who conclude that on average, aggregate aid does not lead to more growth. However, these studies did not settle the dispute. Using a completely different identification strategy based on lagged and differenced data, Clemens *et al.* (2012) find a positive effect of disaggregated aid on growth. They argue that the timing and the distinction between "early-impact" and "late-impact" aid is crucial if its effectiveness is to be tested empirically. More recently, Chauvet and Ehrhart (2014) showed a positive relationship between aid and sales growth at the firm level. Using firm level panel data, they argue that aid increases the productive capacity by relaxing infrastructure constraints, most importantly related to electricity and transport infrastructure.

Finally and most recently, this time back at the macro level, Galiani *et al.* (2014) use an innovative quasi-experimental approach to identify a large positive impact of aid on growth. But due to the specifics of their identification strategy, this result is only obtained by looking at a comparatively small and specific sample consisting of 39 countries. Hence, exemplary for the recent macro literature on aid and growth, Galiani *et al.* conclude that "overall foreign aid increases economic growth among poor countries", but that "aid may have heterogeneous effects depending on recipient characteristics, aid modalities, and donor motives" while they further note that their "relatively small and homogeneous sample is not ideal for testing heterogeneous effects of aid" (p.31).

2.2 Grants *versus* loans

Whereas the heterogenous effect of aid on growth with respect to recipient characteristics and donor motives has been researched extensively, studies that concentrated on the effect of different aid types and modalities have been scarcer.[4] For example, Miquel-Florensa (2007) finds that tied aid, which is aid that has to be used for purchasing goods and services from the donor country, is less effective than untied aid, at least in countries with good policies. As another form of aid modality, much attention has been paid to the concept of conditional aid, which builds on the idea that the actual size of aid flows is made *ex-ante* conditional on the actually realized (*ex-post*) performance (Svensson (2003), Scholl (2009)). Finally, donors have to decide whether to give aid as non-repayable grants or as concessional loans. From a theoretical point of view, both types of aid may have very different effects on growth, which will be discussed in the remainder of this chapter.

2.2.1 Asymmetric information and moral hazard: The case for loans

Foreign aid donors (single donors as well as multilateral aid agencies) and receiving countries are in a classical principal-agent relationship with asymmetric information. The objectives of the donor and those of the recipient do not necessarily coincide. The donor (the principal) wants to see its aid transfers being used as effectively as possible, at least in the short run, because it is accountable to its voters or stakeholders. The direct aid recipient (the agent), which is often the government itself or at least controlled by the government, may want to maximize its own welfare, for example by using part of the aid money to benefit a small well-connected elite, or by using aid partly to buy votes, which are inefficient strategies from an economic point of view. These instances of moral hazard can occur, because the agent is better informed of its own actions than the principal. It is simply too costly for donors to monitor and control the use of every aid dollar in every aid receiving country. As a consequence, if the costs of monitoring and the risk of moral hazard behavior are too high, donors will decide to stop giving aid altogether.[5]

All these considerations seem to speak in favor of loans and against grant aid, because loans arguably provide better incentives in form of harder government budget constraint, leading to more tax effort and fiscal responsibility (Bräutigam (2000). Accordingly, Gupta *et al.* (2003) show that grant aid decreases government revenues while loan aid increases them, supporting the

[4] Doucouliagos and Paldam (2011) found 103 studies looking at aggregated aid, but only 15 which looked at disaggregated aid flows, *i.e.* aid with different modalities. The result of their meta-regression shows a significant positive effect of grant aid, short-term aid and project aid, while technical assistance and multilateral aid are insignificant.

[5] If fungibility is high, this can further reinforce moral hazard problems. Even if budget aid itself is used for its purpose, it may lead to the diversion of other funds, over which the donor has even less control.

hypothesis that grants may be used as substitutes for taxes. Through this channel, grant aid would decrease countries' ability to collect taxes to finance investments in the future, leading to lower growth prospects. Cohen, Jacquet, and Reisen (2007) argue that due to the adverse incentives that grants can have on fiscal discipline, loans should be kept as an important aid instrument. Intuitively, it seems quite obvious that loan aid should be used to finance productive investments, because the loans have to be repaid later. However, the degree to which the repayment obligation incentivizes governments to use aid flow efficiently depends crucially on the myopic nature of political leaders and on the degree of transparency and political accountability in the receiving countries. Nevertheless, *ceteris paribus* grants are more likely to be used for consumption, since they carry no repayment obligation. According to the argumentation about aid effectiveness in the beginning of Section 2.1, this should tip the balance in favor of loans, which can be expected to lead to more growth than grants. However, a precautionary remark should be made here. Whether loans are used differently than grants at all depends on many other factors, including the typical maturity of the loans. If they are very long-term, if the probability of debt forgiveness is deemed high (both can certainly be the case for many development aid loans) or if leaders are simply myopic, then loans may not be perceived different from grants at all. In this case, one would expect no differential impact on growth.

2.2.2 Debt crises and defensive lending: The case for grants

Despite the fact that a reasonable case for loan aid can be made based on the arguments mentioned above, the so-called Meltzer Commission report prominently argued for a shift from loans to grants and a complete cancellation of poor-country debt.[6] Academic backing for this proposal comes from Bulow and Rogoff (2005) who argue that "the increased risk of debt crisis all too often outweighs any gain ordinary citizens might enjoy from [development] loans" (p.393). According to their argument, loan aid allows poor countries' governments to accumulate more debt than what is justified by fundamental growth prospects and more than is supported by domestic political consensus. The accumulation of unsustainable debt through aid in the form of concessional loans can be even exacerbated by what is called "defensive lending". This term describes the incentive of lenders (*i.e.* donors) to always roll over debt and allow the recipient to

[6] The Meltzer commission was a commission of the US congress with the goal to work out proposals for reforming the World Bank and the International Monetary Fund. Sanford (2002, 746-752) summarizes the main arguments for a shift to grants which have been brought forward in this debate.

repay older loans with new loans once the old loans reach their maturity. Defensive lending happens irrespectively of whether or not the former loans have actually been used productively, because the lender is simply reluctant to take losses on his previous loans. This way, the positive incentive effects of concessional loans would be eliminated. However, Cohen Jacquet, and Reisen (2007) found no proof that donors actually behave this way. But overall, the possibility of loan aid leading to the accumulation of large unsustainable debt in developing countries cannot be easily discarded. As high debt levels increase the likelihood of a debt crisis and other distortions, it is also likely to result in a negative effect of loan aid on growth in the long run. [7]

2.2.3 Grants *versus* loans: Does it matter at all?

The preceding discussion showed that the differential effect of loan aid and grant aid on growth is not clear *a priori*. Neither do we know whether to expect any positive effects at all, nor can we say whether the coefficient of grants or loans should be larger. A third hypothesis has been brought forward by Nunnenkamp, Thiele, and Wilfers (2005), who argue that it is very unlikely that grants and loans have any differential impact at all, because they are simply too much alike from the recipients' point of view. In the end, because the growth effects of loan and grant aid are very much ambiguous *a priori*, the answer to the research question becomes an empirical matter. That is the goal for the remainder of this paper.

3. Data and descriptive statistics

With regard to the data, one of the major concerns in studies that estimate growth regressions is the choice between cross-sectional and panel data, and the appropriate period length if panel data is chosen. This question is still unsettled in the literature, but it seems that the most influential growth studies are those relying on cross-sectional instead of panel data (*e.g.* Barro (1991), Mankiw, Romer, and Weil (1992) and Sala-i-Martin (1997) *versus* Islam (1995)). But in the subsample of growth studies which look at the effect of foreign aid on economic growth, the situation seems to be the other way around, as most scholars prefer to use panel data.[8]

[7] See for example Reinhart and Rogoff (2010), who famously argued that growth rates turn negative if external debt exceeds 90% of GDP in emerging market economies.
[8] Out of the studies mentioned in Section 2, for example Burnside and Dollar (2000), Hansen and Tarp (2001), Clemens *et al*. (2012), Chauvet and Ehrhart (2014) and Galiani *et al*. (2014) rely on panel data. Rajan and Subramanian (2008) and Arndt, Jones, and Tarp (2010) are among those studies who use pure cross-sectional data in

Using panel data has many advantages. It increases the number of observations and hence the degrees of freedom, which is an important benefit, because cross-sectional studies are naturally restrained by the limited number of existing countries. Even more important, panel data makes it possible to use panel estimators such as the first differences or the fixed effects estimators, which can effectively remove all unobserved time-invariant country specific effects from the estimation, reducing the inherent endogeneity bias in aid-growth regressions (see Section 4).

The use of panel estimators has also various drawbacks. A main concern, especially with dynamic panel estimators, is that they introduce an additional, artificial source of endogeneity, leading to the so-called Nickel bias (or Dynamic Panel Bias) when initial GDP is controlled for in the regressions (see Section 5.2). However, there are several approaches, such as the Anderson-Hsiao estimator, which have been developed to account for this source of bias. Hence, this paper follows the conclusion of Temple (1999, p. 113) and opts for panel data estimation, as this still increases the possibility of getting unbiased results and being able to interpret relationships in a causal way. Thus, all equations which are estimated in the following sections, are variants of the following stylized growth equation:

$$GROWTH_{it} = ODA\_G_{it-1} + ODA\_L_{it-1} + LGDP_{it-1} + \boldsymbol{X}_{it} + Country_{i} + Period_{t} + \varepsilon_{it} \qquad (1)$$

The panel dataset covers 158 countries (all countries which are members of the World Bank and ever received aid) over a time span from 1960 to 2010 in the full sample. All time-varying variables are computed as annualized arithmetic averages over five-year periods, which results in a maximum number of 10 periods. However, due to lack of data for many of the controls used in the estimation, the actual country and time coverage is considerably lower, as it can be seen in the summary statistics table for the set of controls used in the baseline regression (Table 1).

their preferred specification. However, especially the earlier studies using panel data did not actually apply panel data estimators, thus treating them as a repeated cross-section and estimating them with Ordinary Least Squares (OLS).

Table 1 – Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
GROWTH	539	2.122534	2.84753	-10.2187	16.58168
ODA	539	2.838169	3.560056	-.1245178	24.51311
ODA_G	539	2.582505	3.466089	0	24.46317
ODA_L	539	.255664	.6276455	-2.39276	3.607764
LGDP	539	7.429716	1.241326	4.861017	10.96933
INF	539	2.114528	.9355947	-.2614282	7.993991
DEPTH	539	42.36541	28.75301	6.00598	243.9438
REVO	539	.2039889	.3857931	0	2.6
TRADE	539	79.16474	51.56236	9.33312	400.2004
BUDG	539	-2.369266	4.674645	-41.74083	19.74644
LEXP	539	4.121849	.1684958	3.371642	4.388924

GROWTH- GDP growth rate in %, ODA- bilateral aid as % of GDP, ODA_G- bilateral grant aid as % of GDP, ODA_L- bilateral net loan aid as % of GDP, LGDP- logarithm of initial GDP per capita, INF- logarithm of (1+inflation rate in %), DEPTH- share of broad money M2 over GDP, REVO- number of revolutions, TRADE- (imports + exports) over GDP, BUDG- general government budget surplus as % of GDP, LEXP- logarithm of life expectancy (at birth) in years

There are only 539 observations for 115 countries, hence on average between four and five observations per country.[9] When lagged values and differenced data are considered later, this number decreases even further. The control variables used in the baseline specification are standard in the growth literature and are almost completely taken from the seminal paper of Rajan and Subramanian (2008).[10] They are supposed to capture the institutional environment as well as the impact of policies and geography on growth. The number of revolutions, $REVO_{it}$, as well as the logarithm of life expectancy at birth, $LEXP_{it}$, can be seen as proxies for the institutional environment, which vary only slightly over time. Both are expected to be negatively correlated with growth. Life expectancy could also capture some of the geographical factors that influence a country's growth prospects. The same may be true for the openness variable, $TRADE_{it}$. Countries with favorable geographic conditions, such as access to water and large trading partners nearby,

[9] Of those 541 observations, more than half are for low-income and lower-middle income countries, which are classified according to the World Bank as countries with an average annual income of less than 4,125 USD. However, around 12 % of the observations in the full sample come from high-income non-OECD members with an annual income above 12,745 USD.

[10] The only change, other than the elimination of time-invariant variables, is the substitution of the Sachs and Warner (1995) openness index by an openness index that measures the trade share, computed as the sum of imports and exports divided by GDP. The reason for this change is that it increases the number of observations considerably.

should *ceteris paribus* have a higher trade share and more growth. However, the trade share also reflects policy decisions, as do the budget surplus ($BUDG_{it}$), the ratio of broad money (M2) over GDP ($DEPTH_{it}$), and the inflation rate (INF_{it}). The first three are expected to be positively correlated with growth, while the inflation rate should have a negative coefficient. A more detailed description of the main variables, their definitions and sources is given in the appendix (Table A).

With regard to the summary statistics in Table 1, the aid variables are perhaps most interesting. ODA_{it} denotes the average annualized total bilateral aid flow (grants plus loans) in the respective period, as a percentage of the recipient country's GDP. On average, countries in the sample received around 2.8 % of GDP as foreign aid, which is a sizeable amount. It can also be seen that the lion's share of aid was given as grants (ODA_G_{it}, 2.6% of GDP on average) as opposed to concessional loans (ODA_L_{it}, 0.3% of GDP on average), where a concessional loan is counted as foreign aid if it has a grant element (on its present discounted value) of at least 25%.[11] Loans in this case are net loans, which means that loan repayments (and debt forgiveness) are already deducted from it. All aid data refer to bilateral official development assistance (ODA), which is aid given by those OECD countries who are member of the Development Assistance Committee (DAC).

Because the empirical strategy employed in large parts of the following sections is applicable for bilateral aid flows only[12], only those are considered in the baseline specifications, in order to make the results of the different approaches comparable. Multilateral aid flows are thus not included in the aid variables showed here, but are considered in the robustness tests (Section 6). Because multilateral aid is excluded, the results can only be interpreted as local average treatment effects (LATE). As Deaton (2010) argues, these effects may not be especially policy relevant and may not even be meaningful in some cases. Hedging against this objection, it should be noted already that the inclusion of multilateral aid flows does not lead to significant changes in the estimation results, as it will be further discussed in the robustness test Section 6.

[11] Using a reference interest rate of 10%.
[12] Most importantly, imputed aid flows (multilateral aid flows traced back to their original bilateral source), which have to be added to the bilateral aid flows in the supply-side instrumentation strategy, are not available on a more disaggregated level, which makes this strategy unsuitable for instrumentation of grants and loans.

To complete the descriptive statistics section and to provide a starting point for the discussion in the next chapter, Table 2 shows the pairwise correlation coefficients between growth and the different bilateral aid variables.

Table 2 – Pairwise correlation between aid and growth

Variable	GROWTH	ODA_G	ODA_G_2	ODA_L	ODA_L_2
GROWTH	1				
ODA_G	-0.1778*	1			
ODA_G_2	-0.1357*	0.8837*	1		
ODA_L	-0.0772	0.0612	0.0192	1	
ODA_L_2	-0.1527*	0.1889*	0.0875*	0.7044*	1

Pairwise correlation, * p<0.05, n= 539, ODA_L_2 - square of ODA_L, ODA_G_2- square of ODA_G

Consistent with previous studies, the simple correlation between aid and growth is negative and significant (on the 5%-level), with a correlation coefficient of -0.18 for grants. The correlation between loans and growth is negative as well, but turns out to be statistically insignificant.[13] However, the negative correlation between measures of aid and growth does not mean that aid in its various forms reduces growth. Instead, it is most likely the result of endogeneity bias due to reverse causality and omitted variables. The next section discusses these sources of bias and analyzes the most promising strategies to eliminate it, especially if the goal is to determine the differential impact of grants and loans.

4. Empirical strategy: How to overcome the endogeneity bias?

4.1 The problem

The main obstacle that every study of the aid-growth relationship has to overcome is the problem of the endogeneity of aid. Usually, the coefficient of foreign aid in a simple cross-country OLS growth regression has found to be negative. This does not necessarily mean that aid reduces growth. It rather reflects the simultaneity bias and shows that the *dominant* relationship between aid and growth is negative and runs from growth to aid (Roodman 2008). Countries with adverse shocks to growth receive more aid, and countries that grow unusually fast receive less aid *as a*

[13] Out of the contemporaneous and squared aid terms, only loans fail to pass the threshold of r>|0.1|, which has been proposed by Chatelain and Ralf (2014) to prevent spurious correlation in the presence of "classical surpressors." However, the simple correlation coefficients for the lagged values are all below that threshold, making multicollinearity a serious concern.

percentage of GDP.[14] This leads to a downward bias of the aid coefficient, which can be very substantial. Again, this does not exclude the possibility of a causal relationship from aid to growth. But it shows that it is very challenging to uncover this causal effect and obtain significant and unbiased, not just spurious results, as it is shown for example by Brückner (2013). The research question addressed in this study intensifies this problem, because it essentially asks for the causal effect of aid on growth, conditional on the type of aid (as a concessional loan or a grant). The relationship between the type of aid and growth is expectedly highly endogenous as well, although the sign of the bias could go either way.

From a theoretical point of view, countries with bad growth performances may receive more grant aid (relative to loans) for simple altruistic reasons, *i.e.* because they are more in need of non-repayable funds. From the donor's point of view, it would also make sense to give loans preferably to countries with good expected growth performances, as they are more likely to repay their debt in the future. This would both lead to a downward bias of grant aid relative to loan aid. However, one could also make the opposite argument, arguing that donors want to reward successful reforms in developing countries, preferably with grants, which may lead to an upward bias of grant aid. Even though the first strand of arguments (arguing for a downward bias of grant aid) is conceived to be theoretically stronger, the direction of the relative bias between grants and loans is ultimately an empirical matter.

The following sections will introduce and critically assess the main strategies that have been used to address the endogeneity of aid. Most importantly, it will be analyzed if and how these strategies can be used or extended to deal with the case of two endogenous regressors, being jointly determined and thus showing a potentially high degree of collinearity. The order in which these approaches are presented is somewhat *ad-hoc*, but one may argue that the identification strategies in the beginning are less complex and more efficient, but also more likely to give biased results than those considered in the end.

[14] This mechanical relationship is obvious, but it also holds true that they receive less aid *per capita*.

4.2 The different strategies to solve the endogeneity issue

4.2.1 Lags and differences: Clemens *et al.* (2012)

In choosing an identification strategy, there is often a tradeoff between robustness and efficiency. The most robust instruments are likely to be rather weak, leading to low statistical power. Clemens *et al.* (2012) choose an identification strategy which is simple and efficient, but only accounts for a part of the endogeneity problem. They use lagged values of aid as a regressor and transform all their data into first differences, which leads to the following equation to be estimated:

$$\Delta GROWTH_{it} = \Delta ODA_{it-1} + \Delta ODA\_2_{it-1} + \Delta LGDP_{it-1} + \Delta \boldsymbol{X}_{it} + Period_{t} + \Delta\varepsilon_{it} \qquad (2)$$

Using lagged aid flows as explanatory variables instead of contemporaneous flows has two positive effects. First, it accounts for the fact that aid may influence growth with a time lag, for example because investments need time until they become profitable. Thus, one would expect aid in period *t*-1 to have an impact on growth only in the next period *t*. The second effect is that lagging aid reduces the probability of simultaneity bias. While the growth performance of a recipient country may very well affect its aid inflows in the same period (especially if period averages are taken), it is less likely that it affects aid flows in the *previous* period.[15]

First-differencing the data also aims at reducing a part of the endogeneity bias that stems from omitted variables, by removing all country-specific time-invariant fixed effects (or confounding factors), as it can be seen in equation 2, where the variable $Country_i$ from equation 1 cancelled out. However, first differencing and using lagged aid-flows are not a complete remedy against endogeneity bias. It is still possible that donor A increases its aid to recipient B, because it expects a negative growth shock in the next period. On the other hand, one could also imagine a case where the donor increases aid to recipient C, because a newly elected reform-oriented government will likely lead to a positive growth shock in the next periods. Hence, country-specific time-varying unobserved heterogeneity would still lead to an endogeneity bias, and the direction of the bias is not obvious. For this reason, Clemens *et al.* (2012) and other authors using the same identification strategy (for example Dreher, Minasyan, and Nunnenkamp (2014)) remain very cautious and only interpret their results as a positive *correlation* between aid and

[15] Obviously, this depends on the amount of serial correlation in the error terms.

growth, not necessarily a *causal* relationship.[16] However, even this cautious interpretation has been recently criticized by Roodman (2013, 2014), who replicates Clemens *et al.* and clearly lays out the strong exogeneity assumptions needed for unbiasedness of their results.[17]

4.2.2 Demand-side instruments: Burnside and Dollar (2000)

A more robust way to account for endogenous regressors than lagging and differencing is the use of instrumental variable (IV) estimators such as the two-stages least squares (2SLS) estimator. The idea of IV estimation (of which 2SLS and GMM are special cases) is to find one or more variables, called instruments, which are correlated with the endogenous regressor but are exogenous themselves. In other words, the instruments have to be relevant (highly correlated with the endogenous regressor) and valid (orthogonal, *i.e.* uncorrelated with the error term). They also have to satisfy the exclusion restriction, which is to say that they must not directly affect the dependent variable (only indirectly through the endogenous regressor), and thus can be left out in the final equation.

In their influential study, Burnside and Dollar (2000) chose instruments at the recipient country level, notably the logarithm of its population size, the share of arms imports in total imports and several interactions with their policy variable. There are two problems with this choice of instruments. The policy variable is very likely to be endogenous to growth, since it contains the inflation rate and the budget balance, which are both not orthogonal in a growth regression. Thus, policy-related instruments are probably invalid (Tarp and Hansen 2001). Furthermore, even though the recipient country's population size may not directly influence growth, using it as an instrument does not necessarily satisfy the exclusion restriction. Bazzi and Clemens (2013) show that population size has been used as an instrument for various other variables, many of them being possibly related to growth. Unless those variables are all included in the growth regression (which would probably lead to an endogeneity bias on its own), the population size instrument will be correlated with the error term and thus violate the exclusion restriction. Since it has also been shown that the Burnside and Dollar IV strategy relies almost

[16] Additionally, to increase the likelihood that they actually do capture some causal effects, Clemens *et al.* (2012) exclude all aid flows that can only expected to have an impact on growth in the long run, thus using only „early-impact“ aid in their regressions.
[17] Bazzi and Bhavnani (2014) have responded to some of these criticisms.

completely on population size as an instrument,[18] their IV results are not reliable and very likely to be biased. In general, the main problem with their instrumentation strategy is that they pick instruments at the recipient country level, which makes them unlikely to be exogenous and to satisfy the exclusion restriction. In most cases, recipient level candidate instruments provide additional information about a country's expected growth rate, and excluding them as regressors from the second stage inevitably leads to omitted variable bias. Lessmann and Markwardt (2012) extend the Burnside and Dollar strategy by adding instruments which should reflect the recipients past colonial relationship, for example the population share which has a European first language and several colonial relationship dummies. Both can expected to be correlated with aid, and may well be uncorrelated with growth. But even though this extension increases the instrumentation strength compared to the Burnside and Dollar benchmark, the results would still be biased, as long as the original instruments, which are not orthogonal to growth, are kept.

4.2.3 Supply-side instruments: Rajan and Subramanian (2008)

The state-of-the-art instrumentation strategy for aid picks instruments at the donor (and donor-recipient) level and hence models the supply of aid rather than its demand. Since instrumentation is mainly influenced by variables in the donor country, it can reasonably assumed to be exogenous to growth in the recipient country. Rajan and Subramanian (2008) have been the first to introduce this IV strategy, based on earlier work by Tavares (2003). The idea is to model each bilateral donor-recipient aid flow in a gravity equation, using colonial dummies and the population ratio between the two countries. It is assumed that a donor gives more aid to countries to which it has strong historical and cultural ties (represented by colony dummies) and over which it has a bigger influence (the population ratio). The predicted bilateral aid flows are aggregated over all donor countries and this "zero stage" aggregate is then used as a single instrument for aid in the first stage of the usual 2SLS procedure.

Because the second stage equation has as many excluded instruments as it has endogenous regressors and so is just identified, some specification tests such as the Hansen J-test are not implementable. Arndt, Jones, and Tarp (2010) show that the IV approach of Rajan and Subramanian can be improved in many ways. As in the Burnside and Dollar approach,

[18] Clemens *et al.* (2012) show this by including population size in the second stage regression, which immediately leads to a weak instrumentation problem.

instrumentation strategy mainly relies on the population ratios (POP_d/POP_r). Furthermore, Arndt *et al.* argue that donor-specific colonial variables are not orthogonal to growth and thus should be excluded. For example, former colonies may have different institutions than other countries, which have a direct impact on the country's growth prospects (Acemoglu, Johnson, and Robinson (2001)). However, another remedy for this problem is to substitute the donor-specific colonial dummies by a single colony dummy, leading to prefrred zero stage specification of Arndt *et al.*:

$$ODA_{dr} = COLONY_{dr} + LANGUAGE_{dr} + \ln(POP_d/POP_r) + COLONY_{dr}*\ln(POP_d/POP_r) + Donor_d + \varepsilon_{dr} \qquad (3)$$

As a further extension on Rajan and Subramanian, Arndt *et al.* also use different estimators, such as the Fuller (1977) modified Limited Information Maximum Likelihood (LIML) estimator, which shown to be more efficient than 2SLS if instruments are weak (Hahn, Hausman, and Kuersteiner (2004)). The supply-side IV approach has been criticized among others by Bazzi and Clemens (2013). They show that this identification strategy still relies almost completely on population size, similar to the demand-side approach discussed above. If population size is included as an instrument on its own, the instrument constructed in the zero stage loses all its statistical power.[19] A recent contribution of Dreher, Eichenauer, and Gehring (2014) casts further doubt on the supply-side instrumentation strategy of aid. They argue that this strategy only instruments the "geopolitically motivated" part of foreign aid, as it is based on donor-recipient ties. In their paper, they show that this geopolitically motivated aid is less effective in promoting growth than aid given without these motivations. [20] Thus, the supply-side instruments may not lead to unbiased IV estimates, because they instrument only a "selected part" of the sample of endogenous aid. The resulting coefficients should thus only be interpreted as local average treatment effects. While this may be a concern for general aid effectiveness regressions, it is less worrisome for the approach considered here, since it is mainly focused on the differential impact of grant aid and loan aid. If "geopolitically motivated" aid is as likely to be in the form of grants as in the form of loans, the loan-share coefficient will not be affected by this bias. One could argue that geopolitically motivated aid is much more likely to be in the form of grants than in the form of loans, because grants are more similar to a simple transfer. Thus, the grant aid coefficient

[19] This might change once panel data is used and ideology is introduced as an additional instrument. Section 5 shows whether this is the case.
[20] Dreher *et al.* (2010) reach similar conclusions by looking at multilateral aid flows only.

would be upward-biased. However, the fact that grant aid might be more politically motivated and thus less effective than loan aid can be considered a part of the hypothesis that is tested in this paper. Hence, politically motivated aid supply is not a source of bias, but rather a "channel" through which grant aid could be more or less efficient than loan aid. The conclusion of Dreher, Eichenauer, and Gehring (2014) can be viewed as complementary to rather than competing with the analysis in this paper.

4.2.4 Quasi-experiments : Galiani *et al.* (2014)

Finally, the use of quasi-experiments is a different approach to achieve identification and estimate causal relationships when endogeneity issues are present. The distinctive feature of a quasi-experiment (or natural experiment) is that in most cases it uses just one, clearly exogenous binary shock, to identify a treatment effect.[21] Because units of observation are observed before and after the shock (*i.e.* the treatment), the treatment basically divides those units into a treated and an untreated control group, very similar to actual experiments. But unlike in real experiments or in a randomized controlled trial (RCT), which is becoming more and more popular in economics, the treatment in quasi-experiments is not completely random. The treated group may have unobserved characteristics that make it different from the untreated group, even if no treatment would occur. If this is the case, the treatment assignment is endogenous, leading to a sample selection bias.

With regard to the effects of aid, the treatment or shock that is exploited can be either on the donors or on recipients of aid. Werker, Ahmed, and Cohen (2009) use large oil price shocks as an instrument to identify the effect of foreign aid given from rich Arab countries to poorer Muslim countries. They find no significant impact of aid on growth, but a strong positive effect on consumption and imports. Nunn and Qian (2014) show that US food aid, instrumented by shocks to US wheat production, increased the incidence, onset, and duration of civil war in the receiving countries.

The recent study of Galiani *et al*. (2014) uses a quasi-experimental approach with treatment at the recipient level to identify the effect of aid on growth. The natural experiment they employ

[21] Estimation methods using a quasi-experimental approach are, for example, Differences-in-Differences, Regression Discontinuity Design or Propensity Score matching. Instrumental variables estimation can also be used as a quasi-experiment, as it will be shown below.

as an instrument for aid is the crossing of the income threshold which is used by the International Development Association (IDA) to determine whether a country is eligible for aid. Once the per capita income of a receiving country is above this threshold, IDA aid flows are considerably reduced. Other multilateral aid agencies as well as bilateral donors use the IDA threshold as a signal for their own aid allocation as well. Thus, even though aid from IDA accounts on average for less than 10% of total aid, the crossing of the IDA threshold from below reduces aid inflows by a much larger percentage, due to the herding in aid allocation of donors. In the sample of Galiani *et al.* (2014) which includes 35 countries and the time from 1987-2009, aid decreased by a total of 59% after the threshold was crossed. A dummy which equals 1 when the recipient country's *per capita* income is above the threshold and 0 otherwise should thus be strongly negatively correlated with aid inflows into the same country. Because of that, the "crossing" dummy is a candidate instrument for aid in the first stage of a 2SLS regression of growth on aid. In this case, one can speak of a quasi-experimental approach, because the IDA threshold, which is the cutoff that determines whether a country receives the treatment or not, is not chosen by the researcher or the country itself, and it is also not dependent on individual country characteristics. Instead, it is set exogenously by the World Bank (of which IDA is a member) and it is the same for all countries and over time, except for a yearly inflation adjustment.

However, the fact that the cutoff threshold is determined exogenously does not mean that the crossing of this cutoff is exogenous as well. If the crossing dummy is correlated with the error term in the growth regression, the results will be biased. This may for example be the case if a country crosses the threshold due to a few successive positive growth shocks, after which it will experience negative shocks that bring it back to its balanced growth path. In this case, the reduction of growth rates after the crossing will be falsely attributed to the reduction in aid, even though it only represents a usual reversion to the growth trend. Galiani *et al*. (2014) address this threat to their identification strategy in various ways. Most importantly, they develop a separate growth model and use it to predict the time of the IDA threshold crossing. By using the predicted rather than the actual crossing as an instrument for aid, they ensure that the instrument is not correlated with growth shocks and is thus completely exogenous. With this innovative and convincing approach, Galiani *et al*. (2014) find a comparatively large, significant, and robust positive effect of aid on growth. They further show evidence suggesting that aid increases growth via investment, at least in the short run. However, due to their quasi-experimental approach, they can only include countries that have crossed the IDA threshold from below since 1987, leaving them with a rather homogenous sample of 35 countries. This homogeneity of the sample together with the relatively low number of countries make this approach not ideal for the purpose of this paper. But most importantly, since IDA gives both loans and grants, the crossing of the IDA threshold cannot be used to identify their differential effect on growth.

5. Results

The following section compares the results of several estimation strategies discussed in the previous chapter. To provide a starting point, the model is first estimated as a random effects panel estimator, which does not eliminate country-specific unobserved effects (Section 5.1). Those effects are subsequently eliminated by estimation in first differences (Section 5.2), following Clemens *et al*. (2012). Finally, a supply-side estimation strategy, in the tradition of Tavares (2003) and Rajan and Subramanian (2008), is implemented in an attempt to eliminate any remaining endogeneity bias (Section 5.3). Additional instrumentation and estimation strategies are then considered as robustness test, in Section 6.

5.1 Random effects

Table 3 shows the results of estimation of equation 1 (see Section 3) with the Feasible Generalized Least Squares (FGLS) estimator, thus assuming the Random effects model to be valid. As in all of the following estimations, a variance-covariance matrix which is robust to clustering at the country level is used. Period dummies are employed as well to capture period-specific shocks. Outliers are always identified and excluded according to the BACON algorithm introduced by Billor, Hadi, and Velleman (2000) (see also Weber (2010)). Until noted otherwise, the disaggregated aid variables (loans and grants) and later also their squared values are lagged by one period, to allow the aid flows to "work" and financed investments to materialize returns. Column 1 shows the results when loan aid and grant aid are simultaneously introduced into the growth regression, which implicitly assumes that they have a linear marginal effect on growth.

Table 3 – Random effects

	(1)	(2)	(3)	(4)	(5)	(6)
ESTIMATOR	RE	RE	RE	RE	RE	FE
VARIABLES	Growth	Growth	Growth	Growth	Growth	Growth
L.ODA_L	0.270**	0.246**	0.257	0.075	0.139	0.175
	(0.136)	(0.115)	(0.213)	(0.126)	(0.109)	(0.130)
L.ODA_L_2		-0.041*	-0.046	-0.064**	-0.035	-0.034
		(0.023)	(0.079)	(0.029)	(0.024)	(0.033)
L.ODA_G	0.123***	0.337***	0.246**	0.322**	0.095	0.671***
	(0.044)	(0.116)	(0.114)	(0.143)	(0.111)	(0.144)
L.ODA_G_2		-0.012**	-0.007	-0.013**	-0.004	-0.025***
		(0.005)	(0.005)	(0.007)	(0.005)	(0.007)
LGDP	-0.131	0.040	0.322	-0.169		0.820
	(0.233)	(0.263)	(0.311)	(0.260)		(0.567)
L.LGDP					-1.227***	
					(0.255)	
Observations	539	539	296	329	539	539
Controls	X	X	X	X†	X	X
Period dummies	X	X	X	X	X	X
R-squared	0.268	0.260	0.337	0.407	0.280	0.349
Number of countries	115	115	63	64	115	115

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; X†- time invariant controls added (geography, ethnic fractionalization, ICRG institution index, Sub-Saharan Africa and East Asia dummies); FE-Fixed effects; RE- Random effects; L- value lagged by one period

Both coefficients are positive and statistically significant. A one percentage point increase of loan aid is associated with a 0.27 percentage point increase in the growth rate. For grants, the respective increase equals 0.123 percentage points. The size of both coefficients is close to the typical marginal effect for aggregated aid that can be found in the literature.[22] The linear marginal effect of loans seems to be considerably higher than the effect of grants. However, the Wald test for the equality of coefficients cannot reject the hypothesis that both coefficients are in fact equal (p-value = 0.42). Furthermore, the results in Column 1 are most likely biased. A linear specification of both aid variables may lead to a misspecification bias, if their true effect on growth shows diminishing returns, as it is often found in the literature (see for example Hansen and Tarp, 2001). Hence, in Column 2 I introduce the squared values of loans and grants to account for this possibility. The fact that both squared terms have negative and significant

[22] Clemens *et al.* (2012) find a marginal effect (at the sample average) between 0.1 and 0.2 percentage points, Hansen and Tarp (2001) find approximately 0.1 percentage points with OLS, but a much larger effect with GMM (up to 1 percentage point). Galiani *et al.* (2014), who only consider a log-linear effect of aid on growth, find an increase of around 0.35 percentage points. Finally, Arndt, Jones, and Tarp (2014) find an effect ranging between 0.13 and 0.25 percentage points.

coefficients confirms the expectation that both types of aid seem to work with decreasing marginal returns. Again, both linear terms are positive and highly significant. But similarly to Column 1, the coefficients are not significantly different from each other (Wald test *p*-value = 0.56). The turning point, after which the effect of aid on growth becomes negative, is only around 3 percent of GDP for loan aid, which seems rather small. However, there are only 2 observations in the sample which show a higher average share of loans over GDP, annualized over a five-year period. For grants, the turning point is at 13 percent of GDP. Only 5 percent of the sample observations are above that threshold.

As in Column 1, the overall fit of the model in Column 2, most of the control variables are significant and all but have the expected signs (not shown). Column 3 restricts the sample to countries in the low-income and lower-middle-income group, according to the World Bank. It could be expected that aid works particularly well in these countries, since they face the strongest constraints to external financing sources, which aid could relieve. On the other hand, as they may lack sufficient financial institutions and absorption capacity, loan aid should work considerably less well than grants. Only the second of these assertions is illustrated in Column 3. The coefficients of loans and grants are both lower than with the full country sample, but only grant aid remains statistically significant.

The baseline specification contains only a time-varying set of control variables, because time-invariant regressors have to be omitted in the later steps, when the data are first-differenced. But up to this point, the random effects model allows the estimation of time-varying as well as time-invariant regressor. Thus, Column 4 adds an additional set of standard time-invariant controls, taken from Rajan and Subramanian (2008), to the baseline set of controls. The coefficients of grant aid and squared grant aid in Column 4 remain close to those in Column 2. The coefficient of loans, however, becomes insignificant, once time-invariant regressors are introduced.

This first, simple analysis seems to suggest that grants work better than loans, even though both types of aid have a tendency to be positively related to growth. But there are two main threats to the naïve identification strategy employed in this section. First of all, donor countries may jointly determine the amount of loans and grants that they give in a certain year, which would lead to a high correlation between loan aid and grant aid. Even more important, the aid variables are highly correlated with their squared terms, relative to their correlation with growth (see Table 2). Chatelain and Ralf (2014) show that this multicollinearity problem, originating in the "suppressor variables" (*i.e*. the squared terms), can lead to a situation where identification of the true parameters becomes theoretically impossible. One way to detect multicollinearity is by calculating the Variance Inflation Factor (VIF) for the suspicious regressors. As a rule of thumb, VIFs larger than 10 indicate a severe multicollinearity problem (see for example Wooldridge 2013, p. 98). For the baseline regression of Column 2, this time estimated with OLS instead of random effects, none of the variables has such a high VIF. However, values between 7 and 9 for grants and squared grants indicate that multicollinearity may still be a problem to a certain degree. As an additional test, the so-called condition number is computed. Greene (2012, p.130) regards values higher than 20 as indicative of a multicollinearity problem, while Cameron and Trivedi (2005, p.350) set the threshold at 100. For the current specification, the condition number equals 6.4 and thus indicates no severe multicollinearity problem. However, following the suggestion of Chatelain and Ralf (2014), after which the suspicious "surpressors" may be dropped, I estimate most of the following specifications twice, with and without squared aid terms.

Another potential problem with the simple random effects model arises if unobserved country specific effect are present. The model assumes that those effects (captured by $Country_i$ in equation 1, Section 3) are purely random, *i.e.* completely uncorrelated with the error term ε_{it}. However, in most cases this will not be the case, even with an extensive set of control variables. If there are unobserved effects which are correlated with the error term, the parameter estimates of the random effects model (estimated with FGLS) are biased (see Section 4.1). Hence, column 6 re-estimates the specification of column 2 on the within-transformed data, which eliminates part of the country-specific effects. A Hausman test shows that the models of columns 2 and 6 are significantly different from each other (at the 1%-level), implying that the Random effects estimates are inconsistent. Hence, the next section will use first-differenced data, which also removes the country-specific effects and thus tends to reduce the bias relative to the random effects model.

5.2 First-difference estimator

Table 4 shows the results of the estimation in first differences, following Clemens *et al*. (2012). Compared to the random effects panel data model, first-differencing has the advantage that it eliminates all unobserved (and observed), time invariant country-specific effects, reducing a possible endogeneity bias (see Section 4.2.1). The columns in Table 4 are organized similar to the first three columns in Table 3. First a linear effect of aid is assumed, then squared terms are introduced, and finally the sample is restricted to low-income countries.

The first three columns re-estimate the model of Table 3, now using first-differenced data. The coefficients of loans and grants show a similar pattern as before, although they are much larger now (comparable to the fixed effects version in Table 3, col. 6). Both are significant and positive in the linear specification, and loans seem more effective than grants. When quadratic terms are introduced, the grant coefficient becomes much larger, although both remain positive and significant. In this case, we can also determine that the coefficients of loans and grants are significantly different from each other (the Wald test returns a *p*-value of 0.0102). The turning point for aid is similar to the one found in Table 3, Column 2. Once the country sample is restricted, loans become insignificant. The controls have the expected sign (not shown), with one exception. The initial logarithm of GDP (in its first difference) is highly significant and positive in Columns 1 to 3. From a theoretical point of view, this seems counterintuitive, as it is usually thought of capturing convergence effects. But since the first difference of logarithmic initial GDP is simply the growth rate in period t-1, $LGDP_{it-1}$ is now a lagged dependent variable and its positive coefficient shows that growth rates are positively serially correlated. Unfortunately, the introduction of a lagged dependent variable, leads to the error term being correlated with the dependent variable.[23] The whole estimation becomes inconsistent, and all parameter estimates are potentially biased. This so-called Dynamic Panel or Nickell (1981) bias converges to zero if the number of periods goes to infinity. Hence, it is in particular a problem in relatively short panels.

[23] This becomes obvious by looking at equation 2 and noting that $\Delta LGDP_{it-1} = GROWTH_{it-1}$. This lagged dependent variable is of course correlated with ε_{it-1}, and because this error term is a part of $\Delta\varepsilon_{it}$, $\Delta LGDP_{it-1}$ is endogenous in equation 2.

Table 4a – First difference estimator

	(1)	(2)	(3)	(4)	(5)	(6)
ESTIMATOR	FD	FD	FD	FD	FD	FD
VARIABLES	D.GROWTH	D.GROWTH	D.GROWTH	D.GROWTH	D.GROWTH	D.GROWTH
LD.ODA_L	0.425**	0.346**	0.292	0.404**	0.312**	0.346*
	(0.179)	(0.164)	(0.199)	(0.167)	(0.155)	(0.202)
LD.ODA_L_2		0.025	0.090		0.048	0.080
		(0.040)	(0.076)		(0.035)	(0.081)
LD.ODA_G	0.232***	0.711***	0.521***	0.181**	0.536***	0.431**
	(0.079)	(0.161)	(0.187)	(0.070)	(0.159)	(0.186)
LD.ODA_G_2		-0.026***	-0.017**		-0.020***	-0.015**
		(0.007)	(0.007)		(0.007)	(0.007)
D.LGDP	2.918***	3.327***	3.064***			
	(0.638)	(0.643)	(0.696)			
LGDP				-0.206***	-0.156**	-0.154
				(0.058)	(0.060)	(0.103)
Observations	406	406	221	406	406	221
Controls	X	X	X	X	X	X
Period dummies	X	X	X	X	X	X
R-squared	0.226	0.254	0.312	0.207	0.224	0.289
Number of countries	97	97	54	97	97	54

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; FD- First difference estimator; L- value lagged by one period; D- value in first difference

There are different ways of addressing the Nickell bias. One naïve possibility would be to include only the level of $LGDP_{it-1}$, not its first difference, which has been done in Columns 4 to 6. The coefficient of $LGDP_{it-1}$ (in its level) now has the "right" negative sign and is significant in two out of three cases, while the aid variables are surprisingly robust to this specification change. However, the estimates can still not be expected to be unbiased, because the level of $LGDP_{it}$ is obviously correlated with the differenced error term as well. Introducing the lagged difference of $LGDP_{it-1}$ in Columns 7 to 9, as a proxy for the endogenous first difference, does also not eliminate the bias completely. $LGDP_{it-1}$ is still contained in the lagged difference, and is correlated with the error term ε_{it-1}, contained in $\Delta\varepsilon_{it}$. Furthermore, almost all of the aid variables become insignificant.

If instead the twice-lagged difference is used as a control, the aid regressors gain their significance again (Table 4b, Columns 10 to 12). This may be because the twice-lagged difference of $LGDP_{it-1}$ is not correlated with the error term anymore, hence the dynamic panel bias is completely eliminated. But it may also be because the twice-lagged difference is itself only marginally significant, and not a good proxy of the current ("un-lagged") first difference.

Table 4b – First difference estimator

	(7)	(9)	(9)	(10)	(11)	(12)
ESTIMATOR	FD	FD	FD	FD	FD	FD
VARIABLES	D.GROWTH	D.GROWTH	D.GROWTH	D.GROWTH	D.GROWTH	D.GROWTH
LD.ODA_L	0.158	0.159	0.280*	0.385**	0.252	0.406*
	(0.111)	(0.111)	(0.165)	(0.182)	(0.194)	(0.220)
LD.ODA_L_2	0.036	0.036	0.032		0.068	0.027
	(0.032)	(0.032)	(0.081)		(0.082)	(0.113)
LD.ODA_G	-0.002	-0.002	0.095	0.153*	0.485**	0.540**
	(0.113)	(0.113)	(0.140)	(0.078)	(0.193)	(0.216)
LD.ODA_G_2	0.002	0.002	-0.002		-0.017**	-0.018**
	(0.004)	(0.004)	(0.005)		(0.007)	(0.008)
LD.LGDP	-8.680***	-8.733***	-7.988***			
	(0.748)	(0.764)	(1.498)			
L2D.LGDP				-1.065	-0.847	0.141
				(0.649)	(0.701)	(0.655)
Observations	406	406	221	378	378	209
Controls	X	X	X	X	X	X
Period dummies	X	X	X	X	X	X
R-squared	0.472	0.474	0.467	0.208	0.222	0.305
Number of countries	97	97	54	97	97	54

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; FD- First difference estimator; L- value lagged by one period; L2- value lagged by two periods; D- value in first difference

Hence, one could alternatively consider dropping the $LGDP_{it-1}$ variable completely, which is done in Table 4c, Column 13. In this case, we may trade the Nickell bias for an omitted variable bias. However, the results when $LGDP_{it-1}$ is excluded are very much in line with earlier ones regarding the aid variables. Finally, notwithstanding the previous, rather *ad-hoc* efforts of coping with the Nickell bias, the standard approach is to apply the Anderson-Hsiao (1982) estimator, which uses lagged levels of the dependent variable or twice-lagged differences as excluded instruments for the lagged difference. This approach is considered in columns 14 to 17, which apply Anderson-Hsiao using the Two Stages least Squares (2SLS) estimator.

Table 4c
First differences and Anderson-Hsiao IV

	(13)	(14)	(15)	(16)	(17)
ESTIMATOR	FD	AH	AH	AH	AH
VARIABLES	D.GROWTH	D.GROWTH	D.GROWTH	D.GROWTH	D.GROWTH
LD.ODA_L	0.303*	0.021	-0.257	0.142	-0.090
	(0.156)	(0.438)	(0.443)	(0.362)	(0.505)
LD.ODA_L_2	0.047		0.335**		0.249
	(0.036)		(0.147)		(0.204)
LD.ODA_G	0.588***	-0.214	-1.002	-0.080	-0.528
	(0.157)	(0.215)	(0.632)	(0.207)	(1.077)
LD.ODA_G_2	-0.022***		0.030		0.015
	(0.006)		(0.021)		(0.036)
Observations	406	406	406	406	406
Controls	X	X	X	X	X
Period dummies	X	X	X	X	X
R-squared	0.219	-	-	-	-
Cragg-Donald F-stat	-	23.750	17.901	2.439	1.230
Kleibergen-Paap F-Stat	-	8.495	6.708	1.760	0.899
KP LM Stat (p-value)	-	0.002	0.006	0.199	0.350
Number of countries	97	97	97	97	97
Instrumentation of D.LGDP	excluded	Lagged difference	Lagged difference	Lagged level	Lagged level

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; FD- First difference estimator; AH- Anderson-Hsiao IV estimator, L- value lagged by one period; D- value in first difference;

The results are rather disappointing, as all of the aid variables become insignificant, irrespective of the instrument that is used. However, this failure of the Anderson-Hsiao estimators to return significant results is not specific to the present study, but also shows itself in the original specification of Clemens *et al*. (2012). Once they account for the dynamic panel bias due to the endogeneity of $LGDP_{it-1}$, their positive results vanish.[24] However, there are more efficient, but also more complex GMM estimators which are more suitable for estimating dynamic panel data models, because they can consider further lags and differences as instruments. They will be used in the robustness tests in Section 6.

[24] Understandably they do not further discuss this apparent failure of their identification strategy. I re-estimated Columns 14-17 with aggregated aid and its square term (instead of disaggregated aid), and was still not able to obtain significant results, which illustrates that a large part of the positive effect they find may be due to the Nickel bias.

5.3 Supply-side instrumentation strategy

Table 5 shows the results when loans and grants are instrumented with a "supply-side" IV strategy, in the spirit of Tavares (2003) and Rajan and Subramanian (2008). The different columns show different specifications for the zero stage as well as different approaches to deal with the dynamic panel bias, as all estimations are made with the fixed effects panel IV estimator (2SLS). Using the fixed effects estimator in the second stage is a major difference compared to previous studies, and it should strengthen the instrumentation strategy, because it eliminates unobserved country-specific effects.

The "zero" stage

The specification of the zero stage, which uses exogenous bilateral variables to predict loans and grants at the donor-recipient level, is particularly important (see Section 4.3). In my preferred specification, grants and loans are predicted by dummies indicating a colonial relationship or a common language, the logarithm of the population ratio between donor and recipient and an interaction of the colony dummy with this ratio. The set of predictors in the zero stage thus resembles the one used by Arndt, Jones, and Tarp (2010). However, in my case, there is another crucial point with regard to the zero stage specification. Because two different aid aggregates have to be predicted (grants and loans), it is important to add an additional predictor to the right-hand side of equation 3, which has a significantly different effect on loans than on grants with regard to the donor's choice between these two forms of aid. In a recent contribution, Brech and Potrafke (2014) found a partisan effect on the decision whether to give aid as loans or as grants. More specifically, their results show that left-wing governments tend to give more grant aid than right wing governments, and that this effect is strongest for bilateral aid to low and lower-middle income countries. Because they do not find a significant partisan effect with respect to loan aid and total aid, government ideology seems to be a promising instrument that could capture the donor's decision between grants and loans. Furthermore, it is reasonable to assume that donor ideology is exogenous to growth shocks and other idiosyncratic shocks that could influence the recipient's growth rate.

The instruments from the zero stage are obtained by regressing the bilateral disaggregated aid flows for each donor on all recipients over all periods. The resulting coefficients are then used to predict each donor's "aid flow vector", which contains the predictions of aid flows from this particular donor to each recipient in each period (as a share of the recipients GDP).[25] These vectors are then aggregated over all donors, to obtain a single predicted share of aid over GDP for every recipient-year combination. The whole process is done once for grants and once for loans, which delivers two predicted values ($ODA_0\_G_{\mathrm{it}}$ and $ODA_0\_L_{\mathrm{it}}$) that can then be used as excluded instruments in the first stage of the 2SLS estimation. With regard to the strength of the zero stage regressions, the significance of the different predictors is very heterogenous across donors, but the overall fit of the regressions seems reasonably good. The new ideology index is significant at least at the 5% level in 8 out of 10 cases, if just the 5 major donors are regarded. If left-wing donors give more grants than loans, the coefficient of the ideology index should be larger for grants than for loans.[26] Thus, including ideology may increase the chances of finding a significant difference between grants and loans in an instrumental variables approach.

Second stage results

Table 5 shows the results obtained by following the supply side IV strategy outlined above. Column 1, which uses OLS excludes ideology in the zero stage, shows a significantly positive coefficient for grant aid, while loan aid is not significant.[27] The grant coefficient is very large, about 3 times as large as those found with other identification strategies. However, it is significant at the 5% level. Column 2 presents the results when the random effects panel estimator is used in the zero stage. The coefficients hardly change, and grant aid remains significant. Table 5a provides several measures that are commonly used as specification tests in an IV estimation. Most importantly, they should help to assess whether there is a "weak-instrumentation" problem, which would bias the IV results towards their OLS counterparts (see *e.g*. Bound, Jaeger and Baker (1995), Cameron and Trivedi (2009, pp.194), Greene (2012, pp.289)). One way to check for weak instrumentation is to look at the joint significance of the excluded instruments in the first stage regression(s). As a rule of thumb proposed by Staiger and Stock (1997), a first stage F-statistic below 10 indicates a weak instruments problem. However, in the case of more than one endogenous regressor, the Angrist and Pischke (AP) first stage F-statistic should be looked at instead (Angrist and Pischke (2009), pp. 217-18). For the loan

[25] Here the approach differs substantially from Rajan and Subramanian (2008), which do not regress by donor, but by period.
[26] The ideology index is taken from the Database of Political Institutions (DPI, see Beck et al. 2001). It takes the value 1 if the Chief executive country is classified as right-wing, 2 if it is centrist and 3 if it is left-wing.
[27] Similar to Rajan and Subramanian ((2008), part IV), I start by considering contemporaneous instead of lagged aid flows, contrary to the other estimations so far. However, the results with lagged aid flows are shown in Table 5b.

equation, this statistic equals 7.10 in Column 1 (and 6.70 in the Random effects specification in Column 2), while it is considerably larger for the grant equation. This comparison shows, that there may be a weak instrumentation problem for loans. It also shows that it is not clear whether the OLS or the Random effects zero stage specification lead to stronger instrumentation. Another specification test uses the Cragg-Donald F statistic, which is below 2 in both columns, and decreases slightly when random effects are used in the zero stage. However, the Cragg-Donald F statistic is only valid when the error terms are i.i.d. no robust standard errors are used (see Kleibergen and Paap (2006) and Kleibergen and Schaffer (2007)).

Table 5a – Supply side panel IV

	(1)	(2)	(3)	(4)	(5)	(6)
ESTIMATOR	PANEL IV	PANEL-IV	PANEL IV	PANEL IV	PANEL IV	PANEL IV
VARIABLES	GROWTH	GROWTH	GROWTH	GROWTH	GROWTH	GROWTH
ODA_L	2.463	2.504	2.385	2.517		
	(1.853)	(1.985)	(1.873)	(2.020)		
ODA_G	0.889**	0.892***	0.881**	0.905***		
	(0.349)	(0.339)	2.385	2.517		
LGDP	1.900	1.913	1.874	1.943	0.845	0.910
	(1.270)	(1.278)	(1.263)	(1.297)	(0.972)	(1.018)
L.ODA_L					0.393	0.412
					(0.994)	(1.096)
L.ODA_G					0.361	0.380
					(0.233)	(0.248)
Observations	517	517	517	517	517	517
Controls	X	X	X	X	X	X
Period dummies	X	X	X	X	X	X
R-squared	-0.438	-0.453	-0.411	-0.467	0.328	0.325
Number of country	96	96	96	96	96	96
AP F-stat: Loans	7.10	6.70	7.12	6.67	11.06	10.00
AP F-stat: Grants	13.93	15.72	14.34	14.77	15.60	16.18
Cragg-Donald F-stat	1.398	1.375	1.411	1.361	3.884	3.266
Kleibergen-Paap F-Stat	9.943	11.227	10.231	10.922	8.765	6.428
KP LM Stat (p-value)	0.322	0.298	0.322	0.302	0.048	0.023
First stage:	OLS	RE				
ODA_L	ODA_L_0	ODA_L_0	ODA_L_0	ODA_L_0	ODA_L_0	ODA_L_0
ODA_G	ODA_G_0	ODA_G_0	ODA_G_0	ODA_G_0	ODA_G_0	ODA_G_0
LGDP	-	-	-	-	-	-

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1;, L- value lagged by one period; ODA_L_0 – excluded instrument for loans predicted in the zero stage; ODA_G_0 – excluded instrument for grants predicted in the zero stage

For the case of robust standard errors, the Kleibergen-Paap rank test has been proposed, leading to the Kleibergen-Paap F statistic.[28] According to this statistic, which increases from 9.9 to 11.2, instrumentation has become stronger with the switch from OLS to Random effects. However, a more formal test of weak instrumentation cannot be made, because the critical values computed by Stock and Yogo (2005) were only based on the Cragg-Donald F statistic. A main result of Columns 1 and 2 is that the choice of the zero stage estimator (OLS or Random effects) does not seem to matter too much.

In Columns 3 and 4, the ideology variable is included in the zero stage regressions. The coefficients of most variables hardly change in both the OLS and the Random effects zero stage specification. The introduction of ideology has also just a minor effect on the several specification measures, the general direction of which seems unclear. Hence, the ideology predictor does not seem to have a strong influence on grants, loans, and their difference, at least according to the present specification. However, because it is theoretically attractive and there are no obvious disadvantages of including it (and the results with ideology are in fact slightly better than without in most of the following specifications), the ideology variable will be retained in the zero stage for the following specifications.

Another specification choice that had been implicitly made so far is the use of contemporaneous instead of lagged aid values as regressors, following Rajan and Subramanian (2008, Section IV). Because both aid and growth are measured as period averages and the optimal period length in growth regressions is still a topic of debate, this choice is not theoretically indefensible.[29] Columns 5 and 6 show the results when lagged aid flows are used instead of contemporaneous flows. The coefficients of both loan aid and grant aid are now considerably smaller than before, and have a roughly similar size. They are statistically insignificant at conventional significance levels, even though this is only marginally the case for grant aid (*p*-values of 0.122 and 0.125 respectively). A look at the specification tests such as the AP F-statistic shows that the first stage seems to provide a significantly better fit now than before, especially for loans. The Kleibergen-Paap Lagrange multiplier (LM) statistic, which tests

[28] A disadvantage of the Kleibergen-Paap F statistic is that it cannot be compared with the critical values for weak instruments derived by Stock and Yogo (2005), which are based on the Cragg-Donald F statistic.
[29] For example, an aid shock in the first year of a period could have a stronger effect on growth the following 4 years than it has in the next period (from year 6 to 10). As a matter of fact, switching from contemporaneous to lagged aid makes it just more likely to capture the long-term effect, but less likely to capture the short term effect of aid on growth.

the null hypothesis of underidentification, *i.e.* that the instruments are relevant enough to identify the endogenous regressor, lets us reject this hypothesis at the 5% significance level. Hence, contrary to the specifications with contemporaneous aid, the model now seems to be identified. This leads to two possible interpretations. Firstly, because the "improved" specification delivers insignificant results, one could argue that there is no "true" effect of both loans and grants on growth.

Table 5b – Supply side panel IV

	(7)	(8)	(9)	(10)	(11)
ESTIMATOR	PANEL IV	PANEL IV	PANEL IV	PANEL-IV	PANEL IV
VARIABLES	GROWTH	GROWTH	GROWTH	GROWTH	GROWTH
L.ODA_L			0.146	0.508	
			(2.588)	(2.585)	
L.ODA_G			0.770	0.640	
			(0.506)	(0.404)	
L.log_ODA_L	0.428				0.116
	(0.935)				(2.318)
L.log_ODA_G	0.365*				0.725
	(0.220)				(0.480)
log_ODA_L		2.209			
		(1.772)			
log_ODA_G		0.875***			
		(0.323)			
LGDP	0.008	0.018	-2.254	-2.892	-0.028
	(0.009)	(0.012)	(2.558)	(2.312)	(0.024)
Observations	517	517	514	484	514
Controls	X	X	X	X	X
Period dummies	X	X	X	X	X
R-squared	0.334	-0.275	0.013	0.074	0.075
Number of country	96	96	96	92	96
AP F-stat: Loans	8.90	8.35	9.31	5.05	7.43
AP F-stat: Grants	13.23	15.51	19.96	12.31	17.35
AP F-stat: log(GDP)	-	-	72.08	32.93	78.58
Cragg-Donald F-stat	4.075	1.489	2.830	2.196	2.798
Kleibergen-Paap F-Stat	6.542	11.228	3.945	2.395	2.723
KP LM Stat (p-value)	0.018	0.318	0.017	0.019	0.004
First stage:	OLS	OLS	OLS	OLS	OLS
ODA_L	ODA_L_0	ODA_L_0	ODA_L_0	ODA_L_0	ODA_L_0
ODA_G	ODA_G_0	ODA_G_0	ODA_G_0	ODA_G_0	ODA_G_0
LGDP	-	-	L.LGDP	L2.LGDP	L.LGDP

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; L- value lagged by one period; L2- value lagged by two periods; ODA_L_0 – excluded instrument for loans predicted in the zero stage; ODA_G_0 – excluded instrument for grants predicted in the zero stage

But secondly, one could also argue that this specification simply does not capture the rather short term positive effect of aid on growth, while it is more likely to capture the rather long-term positive effect of loans. Table 5b addresses two other possible sources of misspecification in the supply side IV approach: the possibility of decreasing marginal returns to aid and the dynamic panel bias due to the inclusion of $LGDP_{it-1}$.[30] Columns 7 and 8 account for decreasing returns of both grants and loans (in lagged and contemporaneous form), as they have been found in the previous section. For this purpose, grants (loans) are now measured as the logarithm of one plus the ratio of grants (loans) over GDP. This monotonic transformation is less flexible than introducing squared terms, as it cannot account for increasing returns, but it allows us to account for diminishing returns without having to introduce two more endogenous regressors (see Galiani *et al.* 2014). The logarithm of grant aid is significant at the 5%-level in the contemporaneous, and this time also slightly significant when lagged values are considered. Loans however remain insignificant. Most of the specification tests point to a strengthening of the instrumentation strategy compared to previous specifications, once logs are considered.

Columns 9 to 12 address the dynamic panel bias, similarly to the previous section, by instrumenting the contemporaneous logarithm of GDP with its lagged values, thus implementing the Anderson-Hsiao estimator. In Column 9, the first lag is used for both $LGDP_{it-1}$ and the aid variables. According to the first stage F-statistics, instrumentation remains strong, but the aid coefficients turn out to be insignificant again (*p*-value of 0.128 for grants). Using the second lag of $LGDP_{it-1}$ for instrumentation (Col. 10) and considering again logarithmic transformations (Col. 11) does not lead to significant estimates. Summing up this section, it can be said that a slight majority of the specifications showed a marginally significant positive impact of grant aid on growth However, it was not possible to identify whether the effect is significantly different than the effect of loans, because the effect of loans could not be instrumented in a sufficient way. In general, it became apparent that correction for the Dynamic panel bias seems to be a major issue that drives standard error upwards and thus reduces significance.

5.4 Summary of the main results

Table 6 sums up the results of the previous Sections 5.1 to 5.3 by comparing the preferred specifications of the three different approaches. The first column (Column 2 from table 3) shows

[30] The zero stage in the second part of table 2 is from now on always estimated with OLS and ideology included.

the results of the Random effects estimation (Section 5.1), according to which grants seem more effective than loans, but both are positive and statistically significant.[31]

Table 6- Summary of the main results

Identification strategy	Random effects	First differences		Supply-side IV (2SLS)		
Original column	(2)	(2)	(11)	(3)	(5)	(9)
VARIABLES	GROWTH	D. GROWTH	D. GROWTH	GROWTH	GROWTH	GROWTH
L.ODA_L	0.246**	0.346**	0.252		0.393	0.146
	(0.115)	(0.164)	(0.194)		(0.994)	(2.588)
L.ODA_L_2	-0.041*	0.025	0.068			0.770
	(0.023)	(0.040)	(0.082)			(0.506)
L.ODA_G	0.337***	0.711***	0.485**		0.361	0.770
	(0.116)	(0.161)	(0.193)		(0.233)	(0.506)
L.ODA_G_2	-0.012**	-0.026***	-0.017**			
	(0.005)	(0.007)	(0.007)			
ODA_L				2.385		
				(1.873)		
ODA_G				0.881**		
				(0.346)		
LGDP	0.040	3.327***		1.874	0.845	-2.254
	(0.263)	(0.643)		(1.263)	(0.972)	(2.558)
INF	-0.388**	-0.586**	-0.673**	-0.382	-0.480**	-0.459**
	(0.196)	(0.227)	(0.265)	(0.279)	(0.201)	(0.231)
REVOL	-0.613**	-0.556	-0.624*	-0.943*	-0.775**	-0.919**
	(0.270)	(0.337)	(0.345)	(0.502)	(0.302)	(0.376)
TRADE	0.061**	0.030	0.008	0.098*	0.037	0.012
	(0.030)	(0.045)	(0.044)	(0.057)	(0.027)	(0.059)
BUDG	0.011***	0.019	0.013	0.003	0.016*	0.012
	(0.004)	(0.012)	(0.011)	(0.013)	(0.010)	(0.015)
DEPTH	-0.030***	-0.011	-0.009	-0.048***	-0.040***	-0.018
	(0.008)	(0.013)	(0.015)	(0.016)	(0.010)	(0.023)
LEXP	6.980***	2.078	3.871	8.011	3.621	-0.541
	(1.248)	(2.511)	(3.063)	(5.133)	(2.352)	(5.097)
L2.LGDP			-0.847			
			(0.701)			
Observations	539	406	378	517	517	514
R-squared	0.260	0.254	0.222	-0.411	0.328	0.013
Number of countries	115	97	97	96	96	96
Instrumented variables	-	-	-	ODA	ODA	ODA, LGDP

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

However, their difference turns out to be statistically insignificant and the results are only unbiased under very strong and unreasonable assumptions, including zero correlation between the

[31] Because both the linear and the quadratic term of grants are larger than their loan counterpart, the marginal effect of grant on growth is larger at zero and at any positive level of aid over GDP.

error term and any country-specific, time-varying and time-invariant effects.[32] Columns 2 and 3, which show results from using the First-difference estimator (Section 5.2), eliminate the second of these two assumptions. For unbiasedness, it only has to be assumed that the error term is uncorrelated to unobserved time-varying country-specific effects. Under this weaker assumption, Column 2 still suggests that grants are more effective than loans in promoting growth. Relative to the previous column, the size of the aid coefficients increased rather strongly, and the difference between the two (more exactly the difference of the marginal effect when aid levels are zero) is now significant at the 5%-level. However, those results should not be trusted too much, even if there is no unobserved heterogeneity. The elimination of fixed effects by first-differencing removes one source of endogeneity bias by introducing another, namely the dynamic panel bias. In this case, where the bias stems from the $LGDP_{\mathrm{it-1}}$ control variable, it is very likely that this leads to an upward bias of the aid coefficients (see Roodman 2013). Column 3 shows a specification where this is accounted for by using a twice lagged version of $LGDP_{\mathrm{it-1}}$ as a proxy for its contemporaneous value. Under the main assumptions of serially uncorrelated errors and no unobserved heterogeneity, this could render consistent estimates of the aid coefficients. The grant coefficients remain significant (at the 5%-level) and the linear term reduces to a size which is more consistent with the previous results and those found in the literature. Loans become insignificant, the difference between the two linear terms becomes insignificant as well (F-test *p*-value of 0.22), but the two are jointly significant at the 5%-level, and all four aid terms together are jointly significant as well (at the 10%-level).

The last three columns of Table 6 show results of the supply-side instrumentation strategy using 2SLS (Section 5.3), which aimed to eliminate the last source of bias, the time varying unobserved heterogeneity, by instrumenting grants and loans with exogenous instruments taken at donor-level. Column 4 again shows the instrumented, contemporaneous grant coefficient to be significant at the 10%-level, while (instrumented contemporaneous) loans are insignificant. Both together are jointly significant at the 5%-level, but the difference between the two is not. The size of the grant coefficient is again rather large, implying that a one percentage point increase in the aid to GDP ratio is associated with a 0.88 percentage point increase of GDP growth within the

[32] The first of these assumptions is slightly weakened by the introduction of a set of time-invariant controls in table 3, Column 4. In this case, the error term has to be only uncorrelated with *unobserved* time-invariant effects for unbiasedness. The coefficients of grant aid are very robust to this change.

same 5-year period.[33] If lagged aid flows are used instead of contemporaneous ones, the coefficient of grants becomes insignificant (p-value = 0.12), although its size reduces to a more credible value, as can be seen in column 5. Coping with the dynamic panel bias, which is done in the last column by instrumenting $LGDP_{it-1}$ with its lagged level, does not change the significance of the results, although it improves the quality of the instrumentation.

The conclusions of the main specifications seem ambiguous and reflect the difficulties inherent in any aid-growth regression. The efficient, but suspicious first-difference estimator found grant aid to be effective in the medium term, while the effect of loans was weaker or insignificant. The much more credible IV estimator based on supply-side instruments, which relies on relatively weak assumptions, found grants to be effective as well, but only in the short run. However, once the Dynamic Panel Bias was controlled for in the most rigorous specifications of this section, no significant effect of aid, whether as grants or as loans, could be identified for the short and medium term. Whether these results are robust or whether a different pattern emerges if the dependent variable, the estimator, or the instrumentation strategy are changed will be discussed in Section 6.

6. Robustness tests

This section briefly discusses whether the results obtained so far are robust to three kinds of changes, dealing with some major threats for the validity for the different identification strategies. First, it will be asked whether the results, that have been obtained by using only bilateral aid flows, can be generalized for multilateral aid as well, or whether they just show a local average treatment effect (LATE), which may not necessarily be policy relevant (Deaton (2010)). If the latter is true, then there is also the danger of omitted variable bias due to the exclusion of multilateral aid in the growth regression. Because available data for multilateral aid flows are not sufficiently detailed for the implementation of the supply-side IV strategy of Section 5.3[34], this first robustness test will be made using the first-difference approach of Section 5.2. A second robustness test addresses the concern of weak instrumentation, especially relevant for the results

[33] While the coefficient of grants is relatively large, it is in a similar range as the one obtained with first-differences (same table, second column), suggesting that the inclusion of squared terms is not the main driver of the results (see Roodman (2008)).

[34] The multilateral aid flows have to be traced back to the original donor countries. These „imputed multilateral" aid flows have in fact been calculated by the OECD-DAC, but they do not make a distinction between grants and loans.

obtained in Section 5.3. Using Monte-Carlo simulations, Hahn, Hausman, and Kuersteiner (2004) show that GMM based estimators (such as the "Continuously Updated GMM estimator" (CUE)) and so-called "k-class" IV estimators, such as the "Limited Information Maximum Likelihood estimator" (LIML) and the Fuller (1970) estimator outperform 2SLS in the presence of weak instruments, especially with smaller samples.[35] Finally, a third robustness test will address the problem of Dynamic Panel Bias, which plagues the results of both Section 5.2 and 5.3, in a more sophisticated manner. More specifically, the "Difference-GMM" (Arellano and Bond (1991) and the "System-GMM" (Arellano and Bover (1995), Blundell and Bond (1998)) estimators will be used to account for the endogeneity of the two disaggregated aid variables (loans and grants) as well as endogeneity in the $LGDP_{\text{it-1}}$ control variable, and to test for serial correlation of the residuals, which is crucial for unbiasedness.

Testing for omitted variable bias: Bilateral versus multilateral aid

Table 7 compares the results of selected specifications from Section 2 (Table 4), which used only bilateral aid flows, to the same specifications when multilateral aid flows are added. The overall pattern of the aid variables seems fairly similar, almost all of them retain their sign, most of them have the same significance level and the size of the coefficients is roughly the same. An exception is the linear grant coefficient, which is somewhat smaller when multilateral grants are included. For this reason, if it seems inadequate to generalize the results of Section 5.2 to the level of aggregated aid flows (bilateral and multilateral), at least with regard to the coefficient sizes. With respect to the mere significance of grants and loans, a generalization seems possible and omitted variable bias does not seem to be a major concern for the results obtained in Section 5, although one has to remain cautious. [36] To be conservative, perhaps the results should still only be interpreted as LATEs, as they are not perfectly robust to the inclusion of multilateral aid.

---

[35] The CUE estimator goes back to Hansen, Heaton, and Yaron (1996), while the LIML estimator has already been derived by Anderson and Rubin (1949). Strictly speaking, 2SLS IV and OLS are also k-class estimators with k=1 or k=0 respectively. A very insightful summary of k-class estimators is given in Baum, Schaffer, and Stillman (2007, ch.4).

[36] In addition to the results shown here, all of the equations in Section 5.1 (Random effects) and 5.2 (First differences) have been rerun with a) multilateral aid flows only and b) multilateral grants and loans and their squared terms added as additional regressors and estimated in one regression with the bilateral flows. In the first case, the coefficients of the multilateral aid variables were similar to those of the bilateral ones in the RE models, but lost their significance in most of the first-difference specifications. In the second case, the bilateral aid coefficients hardly changed, and all of the multilateral coefficients were highly insignificant individually, and jointly insignificant at conventional significance levels as well. Thus, the possibility of omitted variable bias due to the exclusion of multilateral aid can be rather confidently rejected. Results are available upon request.

Table 7 – Bilateral versus total aid

VARIABLES	(1) D.GROWTH	(2) D. GROWTH	(3) D. GROWTH	(4) D. GROWTH	(5) D. GROWTH	(6) D. GROWTH
a) Bilateral aid flows only						
LD.ODA_L	0.346**	0.312**	0.158	-0.257	-0.090	0.303*
	(0.164)	(0.155)	(0.111)	(0.443)	(0.505)	(0.156)
LD.ODA_L_2	0.025	0.048	0.036	0.335**	0.249	0.047
	(0.040)	(0.035)	(0.032)	(0.147)	(0.204)	(0.036)
LD.ODA_G	0.711***	0.536***	-0.002	-1.002	-0.528	0.588***
	(0.161)	(0.159)	(0.113)	(0.632)	(1.077)	(0.157)
LD.ODA_G_2	-0.026***	-0.020***	0.002	0.030	0.015	-0.022***
	(0.007)	(0.007)	(0.004)	(0.021)	(0.036)	(0.006)
b) Bilateral + multilateral aid flows						
LD.MULTI_L	0.369**	0.300*	0.097	-0.342	-0.130	0.317*
	(0.159)	(0.162)	(0.099)	(0.483)	(0.548)	(0.162)
LD.MULTI_L_2	-0.033***	-0.025**	-0.009	0.048	0.024	-0.027**
	(0.011)	(0.011)	(0.008)	(0.051)	(0.057)	(0.011)
LD.MULTI_G	0.403***	0.320**	0.003	-0.251	-0.055	0.355***
	(0.139)	(0.139)	(0.092)	(0.401)	(0.472)	(0.135)
LD.MULTI_G_2	-0.009*	-0.008	0.002	0.003	-0.000	-0.008
	(0.005)	(0.005)	(0.003)	(0.012)	(0.012)	(0.005)
Estimator	FD	FD	FD	A-H (diff.)	A-H (level)	FD
LGDP	First difference	Level	Lagged first difference	First difference	First difference	dropped
Respective column in Table 4	2	5	8	14	16	17
Observations	406	406	406	406	406	406

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1, FD refers to the first difference estimator (estimated with OLS), A-H is the Anderson-Hsiao estimator (estimated with 2SLS), controls are the same as in Table XXX and are not shown here

Better inference with weak instruments: GMM-based IV estimators

Table 8 tests the robustness of the results obtained in Section 5.3 by using two different GMM-based k-class estimators instead of the simple 2SLS estimator to re-estimate some of the most rigorous specifications of the supply-side IV approach. More specifically, the Columns 9, 10 and 11 (Table 5), which all used lagged aid flows and instrumented grants, loans and LGDP, are estimated again, this time using the CUE-GMM estimator and the Fuller k-class estimator (with α=1, following Fuller (1977)), both of which have been shown to perform better than 2SLS in finite samples and with the presence of weak instruments (Hahn et al. (2004)).

Table 8 – k-class estimators and continuously updated GMM

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Corresponding column in Table 5	9	9	9	9	9	10	11
VARIABLES	GROWTH	GROWTH	GROWTH	GROWTH	GROWTH	GROWTH	GROWTH
ESTIMATOR	CUE-GMM	Fuller (1)	CUE-GMM	Fuller(1)	Fuller (1)	Fuller (1)	Fuller (1)
L.ODA_L	0.146	0.177	0.146	0.177	0.177	2.667	
	(1.285)	(1.220)	(1.320)	(1.198)	(2.400)	(3.693)	
L.ODA_G	0.770**	0.748**	0.770**	0.748**	0.748†	0.948*	
	(0.334)	(0.323)	(0.350)	(0.330)	(0.473)	(0.511)	
LGDP	-2.254	-2.334	-2.254	-2.334	-2.334	-1.803	-0.028
	(1.899)	(1.841)	(1.975)	(1.872)	(2.401)	(3.177)	(0.022)
L.log_ODA_L							0.140
							(2.152)
L.log_ODA_G							0.707†
							(0.445)
Observations	514	514	514	514	514	484	514
R-squared	0.013	0.028	0.013	0.028	0.028	-0.493	0.083
Number of countries	96	96	96	96	96	92	96
AP F-stat: Loans	8.49	8.49	8.62	8.62	9.29	4.86	7.42
AP F-stat: Grants	16.27	16.27	14.61	14.61	19.87	13.45	17.35
AP F-stat: log(GDP)	95.88	95.88	102.60	102.60	72.22	32.51	78.58
Kleibergen-Paap F-Stat	2.830	2.830	2.863	2.863	3.945	2.467	2.723
KP LM Stat (p-value)	0.003	0.003	0.003	0.003	0.017	0.017	0.004
Standard errors robust to:	-	-	AC	AC	HAC(cluster)	HAC(cluster)	HAC(cluster)

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1, † p<0.12, LGDP is instrumented by its lagged value (and additionally by its second lag in col.6), (lagged) oda_l and (lagged) oda_g are instrumented by their respective (lagged) predicted values from the zero stage, CUE-GMM refers to the Continuously updated GMM estimator, Fuller (1) is the Fuller (1977) k-class estimator with alpha=1, both are implemented with the panel data fixed effects estimator, AC-autocorrelation, HAC(cluster)- arbitrary heteroscedasticity and clustering on country

The first two columns of Table 8 show the results of a re-estimation of column 9 (Table 4), when spherical disturbances (homoscedasticy and nonautocorrelation) are assumed. In this case, the coefficient of grant aid becomes significant at the 5%-level, contrary to the 2SLS estimation in Section 5.3, where it was insignificant even at the 10% level. The coefficient size is higher than what is usually found in the literature, but that a 1 percentage point increase in Aid/GDP would be associated with a 0.7 to 0.8 increase in growth in the following 5-year period seems not completely unreasonable. Comparing columns 1 and 2, it seems that the choice between the CUE-GMM and the Fuller (1) estimator does not matter much, as the results are very similar. The results in columns 3 and 4, which use standard errors robust to arbitrary autocorrelation for inference (and thus “only” need to assume homoscedasticity for unbiasedness) support this conclusion, which is why the CUE-GMM estimator will be dropped from here on. The switch to autocorrelation-robust errors slightly increases standard errors for the grant and loan coefficients, but grants remain significant at the 5%-level. When the assumption of homoscedasticity is relaxed in column 5, and accordingly standard errors robust to heteroscedasticity (accross countries) and arbitrary autocorrelation are used, grants become only “marginally significant” (p-value = 0.116). The grant coefficient regains significance at the 10%-level, when an additional second lag of LGDP is used as an instrument in column 6, in an attempt to further improve instrumentation. In particular, because the number of instruments now exceeds the number of instrumented variables, it becomes possible to run a Sargan-Hansen-test for overidentifying restrictions, which fails to reject the null-hypothesis (“no overidentifying restrictions”) at comfortable significance levels (p-value= 0.2914). However, besides this, it seems that the quality of instrumentation actually decreased relative to the exactly identified case, as can be seen by lower first-stage F-statistics and a lower Kleibergen-Paap F-statistic for the reduced form regression. Thus, the high and significant grant coefficient should be interpreted with caution. Finally, column 7 re-estimates column 11 from Table 4. To account for the possibility of decreasing marginal returns to aid and avoid misspecification bias, logarithmic transformations of the Aid/GDP ratios have been introduced. The grant coefficient becomes marginally significant (*p*-value=0.114). Due to the log-transformation, the presumed marginal effect on growth is now non-linear. At the average Grant Aid/GDP ratio of the respective sample (roughly 2.7 %), a one percentage point increase of Grants/GDP is associated with a 0.23 percentage point increase in growth. This is remarkably close to values found in the literature for total aid (e.g. Clemens *et al.* 2012, Hansen and Tarp (2001)), and it shows that the results of Section 3.2 are reasonably robust to and can even be improved upon by the choice of the estimator.[37]

Correcting the dynamic panel bias: “Difference”- and “System”- GMM

In a final robustness test, the problem of endogeneity due to the Dynamic Panel Bias, which is only superficially addressed in Sections 5.2 and 5.3, will be analyzed more rigorously. For this purpose,” Difference”- and “System”- GMM estimators are used, which enable us to instrument

[37] An improvement which has not been shown here is that the switch from simple 2SLS to CUE-GMM or the Fuller k-class estimator increases the robustness of the first stage regressions, which could for example be seen by the reduction of the critical values of the Stock and Yogo (2005) weak instruments test (Baum, Schaffer and Stilman (2007, p.13)). In this case however, those critical values are not shown because they are only computed for non-robust standard errors and a maximum of two endogenous regressors.

potentially endogenous variables by a set of their own lags and differences. However, the use of those estimators has been criticized as they are vulnerable to endogeneity bias, especially if samples are small and instruments are weak, (Roodman (2009a, 2009b), Bazzi and Clemens (2013)). Nevertheless, Table 9 shows the results of different GMM specifications, using the same basic set of time-varying controls as in Section 5.

Table 9 – Difference and System GMM

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	GROWTH	GROWTH	GROWTH	GROWTH	GROWTH	GROWTH
Estimator	Diff.-GMM	Diff.-GMM	Diff.-GMM	Diff.-GMM	Diff.-GMM	Sys.-GMM
L.ODA_L	-0.166	-0.444	-0.582	-0.231		0.147
	(0.286)	(0.291)	(0.522)	(0.265)		(0.458)
L.ODA_G	0.242**	0.178*	0.248**	0.571*		0.076
	(0.119)	(0.104)	(0.115)	(0.289)		(0.092)
L.ODA_L_2				-0.032		
				(0.092)		
L.ODA_G_2				-0.019*		
				(0.010)		
LGDP	-4.399***	-1.456	-3.394	0.725	-0.033	-0.108
	(1.472)	(1.474)	(2.355)	(2.176)	(0.024)	(0.567)
L.log_ODA_L					-0.486	
					(0.739)	
L.log_ODA_G					0.238*	
					(0.135)	
Observations	430	430	430	430	430	411
F-stat	6.84	5.19	10.95	39.96	8.91	9.10
Number of country	98	98	98	98	98	97
Number of Instruments	100	92	78	112	113	111
AR(1) (p-value)	0.000	0.000	0.001	0.000	0.001	0.001
AR(2) (p-value)	0.096	0.057	0.045	0.030	0.041	0.035
AR(3) (p-value)	0.240	0.597	0.225	0.923	0.188	0.909
Hansen-test (p-value)	0.539	0.419	0.592	0.884	0.367	0.726
Sargan-test (p-value)	0.255	0.959	1.000	0.928	0.999	0.003

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; L- laged value; (1) treats LGDP as predetermined, (2) treats LGDP as endogenous, (3) uses only third lag and deeper as GMM-style IVs

Column 1, which uses "Difference"-GMM, treats all but one of those regressors as strictly exogenous (uncorrelated with future, present and past error terms) and thus uses all their lagged levels (starting at the current period *t*) as IV-style instruments for themselves. The one exception is $LGDP_{it-1}$, which is assumed to be predetermined (uncorrelated with future and current, but potentially correlated with past error terms).[38] The aid variables are assumed to be endogenous, because they are most likely correlated with current error term as well. Hence, only their twice-lagged levels and deeper lags can be used as instruments.

As can be seen in Column 1, this specification returns a grant coefficient which is significant at the 5%-level and has a reasonable size. Loans are negative, but highly insignificant. The Anderson-Bond test for autocorrelation of the differenced residuals shows, as expected, negative serial correlation of order one. However, at least at the 10%-level, we can also reject the null hypothesis of no autocorrelation of order two (*p*-value = 0.096<0.1). Second-order autocorrelation in the differenced residuals, which translates to first-order autocorrelation in levels, means that the twice lagged levels of the endogenous variables and the once-lagged level of the predetermined variables are potentially endogenous and cannot be used as instruments. Column 2 shows thus the same specification which now assumes LGDP to be endogenous as well, allowing only its twice-lagged level as an instrument. The grant coefficient becomes smaller but remains significant on the 10% level. However, the problem of second order autocorrelation actually increased relative to Column 1. Thus, in Column 3 only lags 3 and deeper lags of the aid variables are used as instruments, which would be still exogenous with second order autocorrelation in the differenced residuals.[39] Because there is no third-order serial correlation (*p*-value = 0.22) and both overidentifying restriction tests (the Hansen and the Sargan test) cannot reject the null hypothesis, the estimation in Column 3 seems consistent. The grant coefficient is positive, significant at the 5%-level and has a reasonable size. Column 4 introduces squared terms, and restricts both the predetermined and the endogenous variables to be instrumented from the third lag onwards. As it could be expected when the true relationship between grants and growth is concave, introduction of the squared term (which is negative and significant) increases the coefficient of the linear term, which remains significant at the 10%-level. The difference between grants and loans is statistically significant at the 1%-level. As in the previous columns, there is no evidence of third-order autocorrelation.

[38] The first difference of LGDP, which is a regressor in the difference equation, is the growth rate from the first year of the last period to the first year of the current period and thus practically a lagged dependent variable. It is obviously correlated with error term in the difference equation, where the dependent variable is the difference of growth rates between *t*-1 and *t*. Hence, only the once lagged level (which is the growth rate from *t*-2 to *t*-1) and deeper lags can be used as exogenous instruments for predetermined variables (Roodman 2009a, p.13).

[39] For the endogenous aid terms, which enter already lagged into the difference equation, the number of lags is always reported relative to their initial lag. Hence, lag 3 refers to period t-4.

As another way of allowing for a non-linear effect of aid, Column 5 re-estimates Column 3 using the logarithmic transformations of the aid variables. The specification seems to give good results, including a rejection of third-order autocorrelation. Grant aid is significant at the 10%-level, but the marginal effect on growth is rather small. A one-percentage point increase at the sample mean is associated with a 0.09 percentage point increase in growth. Column 6 finally introduces the "System"-GMM estimator, which adds moment equations in levels to the "Difference" estimator and instruments them with differences. "System"-GMM is expected to improve upon "Difference"-GMM if lagged levels are bad predictors of current differences, for example if the dependent variable is close to a random walk. Even though this is rather unlikely in the present case, for the sake of completeness column 6 still reports the results of "System"-GMM for the specification of Column 3. The grant coefficient becomes insignificant and the Sargan-test, which is not weakened by many instruments, shows that there are overidentifying restrictions biasing the results. Nevertheless, the results using the "Difference"-GMM estimator were promising and make it more likely that the positive effect of grant aid on growth, identified in most of the specifications in this paper, is not a mere reflection of Dynamic Panel Bias or other types of endogeneity.

7. Conclusion

This study extends the aid-growth literature by analyzing the effect of foreign aid, disaggregated into bilateral grants and loans, on growth. The endogeneity of aid is taken into account and several distinct instrumentation strategies are employed in order to obtain meaningful and consistent estimates, while additional attention has been paid to the identification problems stemming from weak instrumentation and multicollinearity in the presence of multiple endogenous regressors. The results have been mixed. It has become apparent that none of the identification strategies developed so far is designed to work robustly in the case of multiple endogenous variables. This makes it particularly difficult to identify a differential effect between two collinear regressors and leads to "jumping" coefficients that frequently change size or sign. Accordingly, this study failed to clearly identify a *significantly different* effect of loans and grants on growth, notwithstanding the theoretical arguments that such an effect might actually exist. However, during the systematic application of different instrumentation strategies on the same sample with the same set of control variables, an interesting pattern evolved. The most basic identification strategies showed positive correlations between growth both grant aid and loan aid. Once the endogeneity bias was accounted for, loans became insignificant. Grants however were consistently, albeit weakly, associated with growth in the majority of specifications. The estimated marginal effect was in most cases larger than previous estimates, found in the literature for aggregated aid. This is consistent with intuition, because the marginal effect of aggregated aid is a weighted average of the effects of its disaggregated components. It could also reflect the fact that disaggregating aid into its bilateral grant and loan components reduces noise in the data, which makes it more likely to identify a significant effect at least for some of the components. But non-significance of loans does not necessarily mean that loans are ineffective in promoting growth. It might also reflect weak instrumentation, which I was not be able to eliminate completely. Nevertheless, the mostly positive, marginally significant effect of bilateral grants on growth was confirmed by various robustness tests, and the results for multilateral aid have been very similar to those for bilateral aid.

However, several caveats apply. Even though different identification strategies have been used, the results should still be interpreted only as local average treatment effects, because in the most rigorous specifications, only bilateral aid flows could be used. Furthermore, the general pattern of the results seems consistent, but they have proven to be sensitive to specification changes nevertheless. Especially the treatment of the Dynamic Panel bias seems to be a crucial point. The more rigorous it is accounted for, the less strong are the results, which should prevent us from interpreting them as a clear causal relationship. Further research on the link between aid and growth should focus on the treatment of Dynamic Panel bias, if good quasi-experiments are not available for the specific research question that is to be addressed. The use of even more disaggregated aid data, such as geo-localized project aid, seems to provide another promising opportunity for further research.

References

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *The American Economic Review, 91*(5), 1369-1401.

Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Anderson, T. W., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, *18*(1), 47-82.

Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 46-63.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277-297.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, *68*(1), 29-51.

Arndt, C., Jones, S., & Tarp, F. (2010). Aid, growth, and development: have we come full circle? *Journal of Globalization and Development*, *1*(2).

Arndt, C., Jones, S., & Tarp, F. (2014). Assessing foreign aid's long run contribution to growth and development. *World Development*, *69*, 6-18.

Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, *106*(2), 407-443.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2007). Enhanced routines for instrumental variables/GMM estimation and testing. *Stata Journal*, *7*(4), 465-506.

Bazzi, S., & Clemens, M. A. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics*, *5*(2), 152-186.

Bazzi, S., & Bhavnani, R. (2014). Response to Roodman "A Replication of 'Counting Chickens When They Hatch'". *Public Finance Review*, 1091142114539751.

Beck, T., Clarke, G., Groff, A., Keefer, P., & Walsh, P. (2001). New tools in comparative political economy: The Database of Political Institutions. *The World Bank Economic Review*, *15*(1), 165-176.

Billor, N., Hadi, A. S., & Velleman, P. F. (2000). BACON: blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis*, *34*(3), 279-298.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*(1), 115-143.

Boone, P. (1996). Politics and the effectiveness of foreign aid. *European Economic Review*, *40*(2), 289-329.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, *90*(430), 443-450.

Bräutigam, D. (2000). *Aid dependence and governance*. Stockholm: Almqvist & Wiksell International.

Brech, V., & Potrafke, N. (2014). Donor ideology and types of foreign aid. *Journal of Comparative Economics, 42*(1), 61-75.

Brückner, M. (2013). On the simultaneity problem in the aid and growth debate. *Journal of Applied Econometrics*, *28*(1), 126-150.
Bulow, J., & Rogoff, K. (2005). Grants versus loans for development banks. *American Economic Review*, 393-397.
Burnside, C., & Dollar, D. (2000). Aid, Policies, and Growth. *American Economic Review*, *90*(4) 847-868.
Cameron, C., & Trivedi, P. (2005): *Microeconometrics. Methods and applications.* Cambridge, UK: Cambridge University Press
Cameron, C., & Trivedi, P. (2009): *Microeconometrics using STATA*. Lakeway Drive, TX: Stata Press Books.
Chatelain, J. B., & Ralf, K. (2014). Spurious regressions and near-multicollinearity, with an application to aid, policies and growth. *Journal of Macroeconomics*, *39*, 85-96.
Chauvet, L. (forthcoming): On the heterogenous impact of aid on growth. A review of the evidence. In *Handbook on the Economics of Foreign Aid*, ed. Mak Arvin and Byron Lew. Cheltenham, UK: Edward Elgar Publishing.
Clemens, M. A., Radelet, S., Bhavnani, R. R., & Bazzi, S. (2012). Counting chickens when they hatch: Timing and the effects of aid on growth. *The Economic Journal*, *122*(561), 590-617.
Cohen, D., Jacquet, P., & Reisen, H. (2007). Loans or grants? *Review of World Economics*, *143*(4), 764-782.
Dalgaard, C. J., Hansen, H., & Tarp, F. (2004). On the empirics of foreign aid and growth. *The Economic Journal*, *114*(496), 191-216.
Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 424-455.
Djankov, S., Montalvo, J. G., & Reynal-Querol, M. (2009). Aid with multiple personalities. *Journal of Comparative Economics*, *37*(2), 217-229.
Doucouliagos, H., & Paldam, M. (2008). Aid effectiveness on growth: A meta study. *European Journal of Political Economy*, *24*(1), 1-24.
Doucouliagos, H., & Paldam, M. (2009). The aid effectiveness literature: The sad results of 40 years of research. *Journal of Economic Surveys*, *23*(3), 433-461.
Doucouliagos, H., & Paldam, M. (2011). The ineffectiveness of development aid on growth: An update. *European Journal of Political Economy*, *27*(2), 399-404.
Dreher, A., Eichenauer, V. Z., & Gehring, K. (2014). Geopolitics, aid and growth. CEPR Discussion Paper No. DP9904
Dreher, A., Klasen, S., Vreeland, J. R., & Werker, E. (2010). The costs of favoritism: Is politically-driven aid less effective? *Center for European Governance and Economic Development Research Discussion paper (97).*
Dreher, A., Minasyan, A., & Nunnenkamp, P. (2014). Government ideology in donor and recipient countries: Does political proximity matter for the effectiveness of aid? *Kiel Working Paper (No. 1870*)
Easterly, W. (2003). Can foreign aid buy growth? *The Journal of Economic Perspectives*, *17*(3), 23-48.
Easterly, W., Levine, R., & Roodman, D. (2004). Aid, Policies, and Growth: Comment. *American Economic Review*, *94*(3), 774-780.
Ehrhart, H., & Chauvet, L. (2014). Aid and Growth Evidence from Firm-level Data. *Université Paris Dauphine Working Paper.*

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica*, 939-953.

Galiani, S., Knack, S., Xu, L. C., & Zou, B. (2014). The effect of aid on growth: Evidence from a quasi-experiment. *World Bank policy research working paper (6865).*

Greene, W. (2012): *Econometric analysis (7e.).* Harlow, UK: Pearson Education.

Guillaumont, P., & Chauvet, L. (2001). Aid and performance: a reassessment. *Journal of Development Studies*, *37*(6), 66-92.

Gupta, S., Clements, B. J., Pivovarsky, A., & Tiongson, E. R. (2003). Foreign Aid and Revenue Response Does the Composition of Aid Matter? *IMF Working paper 03/176*

Hansen, H., & Tarp, F. (2001). Aid and growth regressions. *Journal of Development Economics*, *64*(2), 547-570.

Hahn, J., Hausman, J., & Kuersteiner, G. (2004). Estimation with weak instruments: Accuracy of higher-order bias and MSE approximations. *The Econometrics Journal*, *7*(1), 272-306.

Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite-sample properties of some alternative GMM estimators. *Journal of Business & Economic Statistics*, *14*(3), 262-280.

Islam, N. (1995). Growth empirics: a panel data approach. *The Quarterly Journal of Economics*, 1127-1170.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97-126.

Kleibergen, F., & Schaffer, M. E. (2007). Ranktest: module for testing the rank of a matrix using the Kleibergen-Paap rk statistic. *Statistical software component. Boston College Department of Economics*.

Lessmann, C., & Markwardt, G. (2012). Aid, growth and devolution. *World Development*, *40*(9), 1723-1749.

Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, *107*(2), 407-437.

Miquel-Florensa, J. (2007). Aid effectiveness: a comparison of tied and untied aid. *York University working paper 2007/3*.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 1417-1426.

Nunn, N., & Qian, N. (2014). US food aid and civil conflict. *The American Economic Review*, *104*(6), 1630-1666.

Nunnenkamp, P., Thiele, R., & Wilfer, T. (2005). *Kiel Economic Policy Papers, (4).*

Rajan, R. G., & Subramanian, A. (2008). Aid and growth: What does the cross-country evidence really show? *The Review of economics and Statistics*, *90*(4), 643-665.

Reinhart, C. M., & Rogoff, K. S. (2010). Growth in a Time of Debt. *American Economic Review*, *100*(2), 573-578.

Roodman, D. (2008). Through the looking glass, and what ols found there: on growth, foreign aid, and reverse causality. *Center for Global Development Working Paper*, (137).

Roodman, D. (2009a). How to do xtabond2: An introduction to difference and system GMM in Stata. *Stata Journal*, *9*(1), 86-136.

Roodman, D. (2009b). A note on the theme of too many instruments*. *Oxford Bulletin of Economics and statistics*, *71*(1), 135-158.

Roodman, D. (2013). A Comment on "Counting Chickens When They Hatch." [http://davidroodman.com/david/Chickens%20comment%206.pdf] (accessed June 3rd 2015).

Roodman, D. (2014). A Replication of “Counting Chickens When They Hatch” (Economic Journal 2012). *Public Finance Review*, 1091142114537895.

Sachs, J. D., Warner, A., Åslund, A., & Fischer, S. (1995). Economic reform and the process of global integration. *Brookings papers on economic activity*, 1-118.

Sala-i-Martin, X. X. (1997). I just ran two million regressions. *The American Economic Review*, *87*(2), 178-183.

Sanford, J. E. (2002). World Bank: IDA loans or IDA grants? *World Development*, *30*(5), 741-762.

Scholl, A. (2009). Aid effectiveness and limited enforceable conditionality. *Review of Economic Dynamics*, *12*(2), 377-391.

Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 557-586.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*.

Svensson, J. (2003). Why conditional aid does not work and what can be done about it? *Journal of Development Economics*, *70*(2), 381-402.

Tavares, J. (2003). Does foreign aid corrupt? *Economics Letters*, *79*(1), 99-106.

Temple, J. (1999). The new growth evidence. *Journal of Economic Literature*, *37*(1), 112-156.

Weber, S. (2010). bacon: An effective way to detect outliers in multivariate data using Stata (and Mata). *Stata Journal*, *10*(3), 331.

Werker, E., Ahmed, F. Z., & Cohen, C. (2009). How is foreign aid spent? Evidence from a natural experiment. *American Economic Journal: Macroeconomics*, 225-244.

Wooldridge, J. (2012). *Introductory econometrics: A modern approach (5e.)*. Boston, MA: Cengage Learning.

Appendix

Table A – Variable definition and sources

VARIABLE NAME	VARIABLE DESCRIPTION	SOURCE
GROWTH	Annual growth rate of real GDP per capita	World Development Indicators (WDI) (2015)
ODA_G	Official development assistance, disbursed, bilateral grants, as % of GDP	OECD- Development assistance committee (DAC)
ODA_L	Official development assistance, disbursed, bilateral net loans, as % of GDP	OECD-DAC
LGDP	Logarithm of initial GDP per capita, in 2005 USD	WDI (2015)
INF	Logarithm of (1+ annual CPI inflation rate)	WDI (2015)
DEPTH	Lagged value of money and quasi-money (M2) as % of GDP	WDI (2015)
REVO	Average number of revolutions (including failed attempts)	Arthur S. Banks (2007)
TRADE	Exports + Imports (goods and services) as % of GDP	WDI (2015)
BUDG	Overall Government Budget Balance as % of GDP (Cash surplus as % of GDP if budget balance is missing)	WDI (2015)
LEXP	Logarithm of Life expectancy at birth (in years)	WDI (2015)
GEOGRAPHY	Time invariant composite measure	Originally Bosworth and Collins (2004), obtained via Clemens et al (2012)
ICRG	Composite indicator of Quality of Government	Institutional country risk guide (2013)
ETHN	Ethnic fracionalization index, time invariant	Obtained from Romain Wazciarg's homepage
SSA, EAS	Dummy for Sub-Sahara Africa and East Asia	
IDEOLOGY	Ideological orientation of main execute party, 1=rightwing, 2=centrist, 3=leftwing	Database of Political Institutions (Beck et al. 2001)
MULTI_G	Official development assistance, disbursed, bilateral and multilateral grants, as % of GDP	OECD-DAC
MULTI-L	Official development assistance, disbursed, bilateral and multilateral net loans, as % of GDP	OECD-DAC